

Improving test ordering in general practice : the effects of individual feedback

Citation for published version (APA):

Winkens, R. A. G. (1994). *Improving test ordering in general practice : the effects of individual feedback*. [Doctoral Thesis, Maastricht University]. Thesis Publishers. <https://doi.org/10.26481/dis.19940630rw>

Document status and date:

Published: 01/01/1994

DOI:

[10.26481/dis.19940630rw](https://doi.org/10.26481/dis.19940630rw)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Improving test ordering in general practice

The effects of individual feedback

Improving test ordering in general practice

The effect of individual feedback

Omslag: Guus van Rooy, RL Design

Improving test ordering in general practice

The effects of individual feedback

Proefschrift

ter verkrijging van de graad van doctor
aan de Rijksuniversiteit Limburg te Maastricht,
op gezag van de Rector Magnificus,
Prof. dr H. Philipsen,
volgens het besluit van het College van Dekanen,
in het openbaar te verdedigen op
donderdag 30 juni 1994 om 14.00 uur

door

Ronald Andreas Gerard Winkens
geboren te Heerlen



THESIS PUBLISHERS
AMSTERDAM 1994

Promotores: Prof. dr P. Pop
Prof. dr J.A. Knottnerus
Prof. dr mr R.P.T.M. Grol

Beoordelingscommissie: Prof. dr G.G.M. Essed (voorzitter)
Prof. dr A.F. Casparie (Erasmus Universiteit Rotterdam)
Prof. dr M.P. van Dieijen-Visser
Prof. dr J. de Maeseneer (Universiteit van Gent)
Prof. dr F. Sturmans

De uitgave van dit proefschrift kwam mede tot stand door een bijdrage van:

- . Glaxo BV te Zeist
- . Janssen Pharmaceutica te Tilburg
- . E. Merck Nederland BV te Amsterdam
- . Zorgverzekeraar VGZ



Contents

1	General "Men more frequently need to be reminded than informed"	
1.1	Context and overview	
1.2	Literature on feedback	
	Study question	
1.3	An outline of the study design	10
2	A diagnostic advice for general practitioners	15
3	Effect of feedback on test ordering behaviour of general practitioners	25
4	A randomized controlled trial of routine individual feedback on diagnostic test requests	35
5	Long-term effects of routine individual feedback on diagnostic actions	45
6	General practitioners and diagnostic testing. Why are there differences in test ordering patterns and in the change of ordering by feedback?	55
7	Does a reduction in test use lead to more referrals?	65
8	Routine individual feedback on diagnostic test requests: an economic evaluation	75
9	Fitting a routine health care activity into a randomized trial	85
10	General Discussion	125
11	Summary	135
	Summarizing	145

Contents

1	General introduction	1
1.1	Context and outlines	1
1.2	Literature on feedback	7
	Study questions	1
1.3	An outline of the study design	16
2	A diagnostic centre for general practitioners	19
3	Effect of feedback on test ordering behaviour of general practitioners	27
4	A randomized controlled trial of routine individual feedback on diagnostic test requests	41
5	Long-term effects of routine individual feedback on diagnostic actions	57
6	General practitioners and diagnostic testing. Why are there differences in test ordering patterns and in the change of ordering by feedback?	71
7	Does a reduction in test use lead to more referrals?	85
8	Routine individual feedback on diagnostic test requests - an economic evaluation -	97
9	Fitting a routine health care activity into a randomized trial	109
10	General discussion	125
11	Summary	137
	Samenvatting	143

Dankwoord	151
Curriculum Vitae	154
Appendices	
I The request form of the Diagnostic Coordinating Centre Maastricht	155
II Two examples of feedback reports	156
III Guidelines used in the Maastricht region	164
IV An example of the guidelines	165
V The costs of feedback	168

General introduction

1.1 Context and outlines

In this book the effects are described of individual feedback on diagnostic test ordering behaviour of general practitioners as provided by the Diagnostic Coordinating Centre Maastricht.

Before describing why the study was performed with this specific design, it is for clarity reasons necessary to sketch the currents in diagnostic testing as well as the development of the Diagnostic Coordinating Centre Maastricht and its feedback.

The trends in diagnostic testing

In medicine, diagnostic tests can be valuable tools in the work-up of a variety of clinical problems. In both primary and secondary health care more and more tests are available to the individual doctor. For many years the use of diagnostic tests has been growing rapidly, but there is doubt about the appropriateness of test use.¹⁻³ Especially in the western countries governments and health care providers are facing an explosive growth in the use of diagnostic resources, which calls for restrictions to achieve a more appropriate test use. Also in the Netherlands there is a growing awareness that the use of diagnostic test facilities should be reduced.⁴ Many tests are ordered simply due to technological imperative. Merely its availability leads to the use of a test, regardless of the fact whether it is appropriate or not.⁵ Studies have also shown that restrictions in the availability, for example by reducing the pre-printed possibilities of tests on request forms, can be effective.⁶

For a majority of tests there is insufficient knowledge on reliability, validity and diagnostic value. As a result, the number of studies on the validity and/or diagnostic value of tests is increasing rapidly.

Especially in those countries (such as the Netherlands) in which primary care plays a major role in the health care system, a substantial number of studies are performed in general practice.⁷⁻¹⁰

More knowledge on the validity and diagnostic value of tests is a first step to achieve a more appropriate test use. Second, this knowledge must be passed on to those who actually need it, namely the doctors working in day-to-day health care. Only then can a behavioural change, like a more appropriate test use, be achieved.

There are several ways for transferring information. Continuing medical education and literature are among the oldest, but in recent decades there is evidence that they are not as effective as is generally assumed.¹¹

Nowadays peer review, audit and feedback are becoming increasingly popular ways to provide information on the validity and/or diagnostic value of tests in order to change the behaviour of doctors, both in primary and secondary care. The Health Council of the Dutch government argues in favour of continuing medical education and/or peer review based on guidelines or standards, like those developed by the Dutch College of General Practitioners, the NHG.^{12,13}

There is growing evidence that feedback is effective, not only in reducing the use of tests but also in improving the performance in daily practice, such as the sampling technique of cervical smears.^{14,15} In the Maastricht region it was attempted to improve the appropriateness or rationality of tests ordered, using individual feedback. The feedback is based on (regional) guidelines and NHG standards as well. This feedback is the theme of this book.

The Diagnostic Coordinating Centre Maastricht

The Diagnostic Centre Maastricht started in 1979 as an experiment initiated by the Ministry of Education and Science. The aim was to improve the relationship, and to serve as a bridge, between primary and secondary health care.¹⁶ The experiment was stopped in 1983 but it was at the request of the general practitioners in the Maastricht region that the Diagnostic Centre Maastricht was continued and its activities were expanded.

In 1987 the Diagnostic Coordinating Centre Maastricht became a foundation, while the general practitioners were more involved in the policies of the centre.

The Diagnostic Coordinating Centre Maastricht processes all diagnostic requests of all approximately 85 general practitioners in Maastricht and its surrounding area. In this catchment area 187000 people are currently living. On an annual base 30000-35000 patients are "referred" to the centre.

A diversity of tests can be requested. They are put out to the laboratories of the University Hospital Maastricht, where the Diagnostic Coordinating Centre

Maastricht is located. Tests from the following categories can be requested: clinical chemistry, haematology, serology, urine and faeces tests, bacteriology, virology, histology, radiology, ultrasonography, electrocardiography, endoscopy and clinical pharmacology. For all requests a specially designed request form is to be used (see appendix I).

A general description of feedback

Feedback is described as a way to "control a system by reinserting into the system the results of its performance".¹⁷ Feedback can demonstrate to a subject the impact of his behaviour, to enable him to make choices regarding his future behaviour.¹⁸ This is also applicable to medical practice.

When feedback is used in medical practice, information on behaviour is provided to the doctor. This can occur during or after the performance that is involved. The feedback from the Diagnostic Coordinating Centre Maastricht is provided afterwards. For the benefit of the feedback request forms from one month are collected and discussed, approximately one or two months later.

The development of the feedback used in the Maastricht region

During the experimental phase of the Diagnostic Centre Maastricht (1979-1983) guidelines for a variety of diagnostic problems were compiled by specialists together with expert general practitioners.¹⁹ After authorization they were distributed to the general practitioners in the Maastricht region.

These guidelines reflected the optimal diagnostic work-up, based on the most recent scientific knowledge and insights. Later in that same period the Diagnostic Centre Maastricht started providing feedback on test requests based on these guidelines. Initially the feedback was on group level. A questionnaire, completed in 1984 showed that the feedback was appreciated, although a more profound and individual feedback would have had more approval.^{20,21} Therefore, in spring 1985, the Diagnostic Centre Maastricht started providing individual feedback.

A few years later the first, but yet considerable, potential effects became apparent which called for further investigation. This was the immediate cause for the study presented in this book. Several aspects of the feedback and closely related matters are studied. In the following chapters these will be discussed.

The contents of this book

First of all, the current state of the art on feedback is studied. What can feedback achieve? What were the results of other studies on feedback as a tool to change the behaviour of physicians? The results of a literature survey are given briefly in the second part of this chapter.

Before the results of the study are presented, the study design is summarized in the third part of this chapter. In chapter 2 the feedback procedure itself is elaborated separately.

The results of the study are presented in the chapters 3 to 8. Before the onset of the study, the feedback had already been in action for approximately 4 years. Therefore, if the feedback was able to change test ordering behaviour, (part of) the effects might have already been achieved before the start of the study. The first effects thus could only be studied retrospectively. To that end, our data were compared to those of a reference laboratory in a retrospective analysis. Its results are presented in chapter 3.

The actual study to demonstrate a causal relation between feedback and a change in test ordering behaviour was a randomized controlled trial. Before that trial only a chronological relation was found between a decrease in the number of requests and the feedback. In the trial we considered the influence of feedback on the volume and quality of requests. The outcome of the trial is presented in chapter 4.

When feedback is capable of realizing a more efficient use of diagnostic tests, the question is how long this change can persist. By the end of our study, feedback had been provided for more than 7 years and therefore long-term effects could be studied. These long-term effects are presented in chapter 5.

Not all general practitioners respond to the feedback to the same extent. To discover the factors that can explain why a general practitioner does, or does not, respond to the feedback, request data were compared with data from a questionnaire. The most important findings are presented in chapter 6.

A more efficient use of diagnostic facilities can theoretically be nullified by a higher referral rate as an adverse side effect of the feedback. The results of a study on the relation between test ordering behaviour and referral behaviour are presented in chapter 7.

In chapter 8 the economic consequences of the feedback are discussed. Although feedback can bring along a decrease in the volume of requests, the feedback procedure itself is time-consuming and therefore money-consuming. We have tried to estimate whether the feedback was economically worthwhile. The reason of the Diagnostic Coordinating Centre Maastricht for providing feedback is to support

the general practitioners in their diagnostic work-up and thus to improve the quality of test ordering behaviour. For others, however, the financial benefits are the major reason for giving feedback.

The design of the trial which was performed is derived from the classical randomized clinical trial, however, with marked differences. The consequences of these adjustments are discussed in chapter 9.

Chapter 10 finally contains a general discussion on the methodology and the results of the study as a whole, including the conclusions that can be drawn. Also in this chapter, recommendations are offered for the implementation of feedback elsewhere and for future research on quality assurance in diagnostic testing.

The final part of the book contains a summary (in English as well as in Dutch) (chapter 11), an acknowledgement (dankwoord) (in Dutch), a curriculum vitae (in English) and several appendices (in English).

Several chapters of this book are reprints of manuscripts that have either already been published or been submitted for publication. Manuscripts are offered to several journals, in different countries. Unfortunately, this shuts the door on a fully consistent use of words. For example, general practitioners are also mentioned as family physicians, and behaviour is also written as behavior.

Purely for stylistic reasons, general practitioners are indicated as males.

References

1. Pinckney ER. The accuracy and significance of medical testing. *Arch Intern Med* 1983; 143: 512-4.
2. Wong ET, Lincoln TL. Ready! Fire!...Aim! An inquiry into laboratory test ordering. *JAMA* 1983; 250: 2510-3.
3. Griner PF, Glaser RJ. Misuse of laboratory tests and diagnostic procedures. *N Engl J Med* 1982; 307: 1336-9.
4. Dunning AJ. Groei in tal en last. Over diagnostische en therapeutische verrichtingen in de gezondheidszorg. Den Haag, 1986.
5. Mandell HN. Technological imperative. Or, when your tool is a hammer, everything looks like a nail. *Postgrad Med* 1983; 74: 25-6.
6. Zaat JOM. De macht der gewoonte. Over de huisarts en zijn laboratorium-onderzoek. Amsterdam: Thesis, 1991.
7. Dinant GJ. Diagnostic value of the erythrocyte sedimentation rate in general practice. Maastricht: Thesis, 1991.
8. Fijten GH. Rectal bleeding, a danger signal? Maastricht: Thesis, 1993.

9. Baselier PJAM. Acute bacteriële urineweginfecties in de huisartspraktijk. Nijmegen: Thesis, 1983.
10. Knottnerus JA. Interpretatie van diagnostische gegevens. Maastricht: Thesis, 1986.
11. Sprij B, Casparie AF, Grol RPTM. Interventiemethoden om een gedragsverandering in de medische praktijkvoering te bewerkstelligen; wat is effectief? Ned Tijdschr Geneeskd 1989; 133: 1115-8.
12. Dunning AJ. Kiezen en delen. Advies in hoofdzaken van de commissie keuzen in de zorg. Den Haag, 1992.
13. Rutten GEHM, Thomas S. NHG-Standaarden voor de huisarts. 1st ed. Utrecht: Bunge, 1993.
14. Axt-Adam P, Van der Wouden JC, van der Does E. Influencing behavior of physicians ordering laboratory tests: a literature study. Med Care 1993; 31: 784-94.
15. Buntinx F. De afnamekwaliteit van cervixuitstrijkjes. Maastricht: Thesis, 1992.
16. Brouwer W, Kerkhof PD, Pop P. De relatie 1e-2e lijn in ontwikkeling. Eindrapport experiment Diagnostisch Centrum Maastricht. Maastricht, 1983.
17. Ende J. Feedback in clinical medical education. JAMA 1983; 250: 777-81.
18. Joyce L. Giving Feedback in Formative Evaluation: A Non-directive Strategy. In: Conner RF, Hendricks M (Eds). International Innovations in Evaluation Methodology. San Francisco: Jossey-Bass, 1989.
19. Beusmans GHMI, Does RJMM. Werkafspraken in het Diagnostisch Centrum Maastricht. Huisarts Wet 1988; 31: 154-8.
20. Pop P, Beusmans GHMI, Knottnerus JA. Een diagnostisch centrum voor Maastricht. Versterking van de eerste lijn en samenwerking huisarts-specialist. Med Contact 1987; 42: 845-9.
21. Beusmans GHMI. Huisarts en diagnostisch centrum. Maastricht: Thesis, 1986.

1.2 Literature on feedback

In this section a brief outline is given of the state of art concerning feedback. Detailed overviews of literature can be found in the many reviews that have already been published.¹⁻⁷

What is known about feedback and its effects? Under what conditions can feedback have its optimal efficiency? Is feedback economically worthwhile? How long can feedback be effective? Who needs feedback and who is responding the best?

Feedback is becoming increasingly popular as a tool to change the behaviour of doctors. This interest started after 1970 and increased gradually. For a look at literature on feedback, therefore it is sufficient to look over the period 1970 until now. Initially we looked at literature from 1983-1992, by means of CD-ROM, using keywords such as feedback, audit, peer review and diagnostic tests. Using references in these publications, other relevant publications were traced, especially those published before 1983.

In general, feedback is provided on a variety of aspects of medical care, diagnostic and therapeutic actions in particular. The publications presented in this survey are solely focused on diagnostic actions.

The number of publications focusing on feedback on test ordering behaviour is very large. This survey therefore will give only typical examples that show the developments within the field of research on feedback. With regard to effects of feedback, only those studies are mentioned in this survey that have included a concurrent control group. Without a suitable control group it is difficult to see what test ordering behaviour would have been without intervention.

The majority of papers were published in the eighties. From 1980, over the years more and more randomized controlled trials were published, although they are still in the minority. All studies discussed diagnostic behaviour in different ways, as described below.

The appropriateness of tests

A reduction in the number of unnecessary requests can be achieved by providing information on the indications for requests. In several studies guidelines or protocols were provided and the effects on the volume of test requests were

determined. The effects of discussing the necessity of test requests in relation to guidelines or protocols seem to be more pronounced than the effects of feedback on test use or costs of tests.⁸⁻¹¹

The studies mentioned so far concerned only quantitative aspects of diagnostic testing. There are also studies in which feedback was used to improve quality of test ordering behaviour. Most of these studies aimed at an increased agreement with previously established standards for good quality of care. Mostly such an increased agreement was indeed achieved.¹²⁻¹⁴

The number of tests

Several studies aimed at reducing the number of diagnostic tests. This type of feedback appears to be effective although the effects are variable.^{9,15-23} Most studies led to considerable changes, except the study of Wilson et al. In this study, only internists showed a small but significant reduction, while surgeons did not change their request habits.²³ In the other studies changes varied from 10% to more than 70%. The wide range might be explained by differences in study set-up, target groups, profoundness of the intervention, the tests that were discussed and many other factors.

Several studies have focused on the number of requests, as well as on the costs of requests.^{17-20,24} When this was done for separate study groups, the results showed that discussing the number of requests need not be more effective than price information.¹⁶⁻¹⁸

The costs of tests

In this type of feedback only price information is provided without discussing test usage and without giving suggestions for improvement. Price information appears to be needed more than one might expect. Prices are generally estimated incorrectly by physicians.²⁵ Although feedback on costs is simple and not laborious, the effects are not clear.^{8-12,24,26,27} Only a few studies showed a clearly positive effect.²⁸⁻³⁰ In only two studies was the test ordering behaviour considered after the intervention was stopped. In the follow-up (one month) Cohen found an even stronger effect of price information on X-ray requests.²⁷ Tierney saw the effects that were achieved in his trial disappear within 5 months.²⁸

Recapitulation

Summarizing all these studies on different forms of feedback, it appears that feedback is capable of changing diagnostic behaviour in different ways. The studies vary largely with regard to the kind of feedback, the procedure that is to be improved and the setting in which the feedback is applied. In a few studies a combination of several strategies is used, making it impossible to trace the effect of the individual strategy. As a result a comparison of the different studies or the different ways in which feedback is provided is practically impossible. Also, no conclusion is permissible on what type of feedback is the most effective or efficient. It appears that more, properly conducted, comparative studies are needed. The methodology of the studies presented in this survey does not always allow valid conclusions. In several studies patients are randomized instead of the physicians. Only few studies are randomized controlled trials. In most studies physicians are asked to participate. This bears the risk of positive selection: especially those physicians participate who have a positive attitude towards feedback. "Voluntary mechanisms are most used by those who least need them".³¹ Apart from the effectiveness of feedback, there are several other aspects that need to be considered next.

The duration of the effects

In virtually all studies feedback was provided for the purpose of that special occasion. When the study was stopped, the feedback was stopped as well. The duration of the intervention varied widely: sometimes feedback was provided only once, but in most studies the intervention period lasted several months with a maximum of 2 years. Only on occasion was the feedback provided for a period of 3 to 4 years.

When effects were achieved, they persisted during the study period. In one study request data were followed over a period of 7 years but several educational strategies were used simultaneously during all 7 years and there was no concurrent control group for the whole period.³² Therefore, the long-term effects of feedback cannot be traced. In most studies the duration of the effects was not determined once the feedback was stopped. In only two studies was the persistence of the effects determined in a follow-up. The effects diminished soon after the feedback was stopped.^{22,28}

Conditions to optimize effectiveness

Apart from the subject of the feedback (costs, volume or appropriateness) other features of the feedback can influence the effectiveness of the feedback. Although the large differences between all studies allow no valid conclusions, the following factors could be important.

From all studies discussed in this survey it seems that the more profoundly the feedback discusses the behaviour of the doctor, the stronger the effect is and the longer it persists. Also, a personal approach seems to be beneficial and preferably the feedback is provided by a respected, expert peer.^{19,22} In those cases where diagnostic behaviour is compared with standards or guidelines, it seems important that these standards or guidelines are accepted and approved by the professional group to which those who receive the feedback belong.³³ It does, however, not automatically lead to a higher response to the feedback.¹¹

The target group for feedback

Feedback has been provided to several professional groups. Especially the American studies have focused on specialists in inpatient and outpatient clinics. European studies have focused more frequently on primary care. It seems that both primary and secondary care doctors respond equally to feedback. So far, sufficient study has not been undertaken into why some doctors respond to the feedback and others do not. Obviously, feedback is only effective when it is provided to the appropriate person: feedback must be presented to those who actually ordered the tests.³⁴ Doctor and practice characteristics are only studied to explain differences in test ordering behaviour itself and not with regard to the willingness to change.³⁵ Factors like education, experience, attitude to risk-taking and many others are considered to correlate with test ordering behaviour.³⁶⁻³⁸ Schroeder found no characteristics identifying high or low users of diagnostic tests.²⁶

Another important issue is how doctors look at the feedback. Feedback can only be effective when it is approved and accepted by the target group. Concerning the attitude towards feedback Metcalfe stated that young doctors show a greater enthusiasm for audit than established doctors.³⁹ Apparently there is still much uncertainty and at times controversy about factors that explain differences in test use and the willingness to respond to feedback.

Side effects of feedback

In all studies little attention has been paid to negative effects or side effects of feedback. The question remains unanswered whether feedback has, aside from the positive effects on test use, undesirable side effects.

Requesting fewer tests can theoretically lead to an underuse of tests, especially if the restriction is done unselectively. Perhaps tests are not requested in situations where they should have been. Unfortunately, this phenomenon is difficult to trace and has therefore hardly been studied so far. Few studies have investigated the occurrence of underutilization of diagnostic tests.

Kroenke found no such side effect.⁴⁰ The risk of an underuse of tests can undoubtedly be minimized by discouraging test use for selected indications as is the case when feedback is aimed at a greater agreement with guidelines or standards for good quality of care.

Another possible side effect is the effect on referral rates. When a general practitioner requests fewer tests, he or she might feel less certain. This might provoke a higher referral rate, with more unnecessary referrals. Literature, however, contains insufficient information on this subject.

The economic effects of feedback

No matter for what purpose the feedback on diagnostic testing is provided, a change mostly means a reduction in the volume. This reduction is accompanied with savings in costs. This is, however, not the only factor determining whether feedback is cost effective. Factors like patient outcome and adverse side effects, costs of feedback itself, etcetera, are important as well.⁴¹ However, like adverse side effects, patient outcome is difficult to study considering the diversity of diseases (with different morbidity, mortality, prevalence, etc.) involved in diagnostic testing. Tierney found no negative effects on patient outcome, meaning no higher number of hospitalizations or outpatient visits.²⁸ Only on occasion are the costs of the feedback itself considered. The outcomes of these studies are contradictory. Some conclude that costs of feedback do not outweigh the savings, while others conclude that the savings can be much higher.^{12,26}

Conclusions

Despite the great number of studies, many questions with regard to feedback remain. The diversity of interventions and settings, together with the complexity of the process of changing doctors' habits make it difficult to obtain comparable and generalizable results. Nevertheless, feedback appears to have an effect on test ordering behaviour, especially with regard to the number of tests ordered.

The extent of the effect is, considering the differences mentioned above, difficult to determine. Routine individual feedback on quantity and quality of requests as provided by the Diagnostic Coordinating Centre Maastricht has not been studied until now. Judging by what literature reveals it appears that several factors that are important for a correct interpretation of feedback and its effects, have been insufficiently studied until now and therefore need more attention.

The effects of feedback on rationality need to be evaluated more thoroughly. Feedback itself should ideally be focused on the volume of requests as well as on the rationality.

Although the potential effects of feedback have been demonstrated in several studies, there is insufficient knowledge as to how long an effect persists. Longitudinal research is needed. It is not possible to conclude what the best way to change test ordering behaviour is and how feedback should be provided to achieve the highest efficiency. There is much need for comparative studies.

Characteristics of the doctor that can predict how he or she will respond to the feedback are largely unknown. Also it is unknown whether characteristics that correlate with test use can predict the response to feedback.

The financial consequences of feedback have been studied only occasionally. In fact they should be determined for every intervention with feedback.

When feedback is applied in general practice, it bears the risk of a higher referral rate or adverse effects on patient outcome. In literature this has been discussed unsatisfactorily so far. Both factors need further attention. Side effects of feedback on patient outcome in particular are extremely difficult to study.

The study, presented in this book, addressed the following questions:

- Are reductions in volume or improvements in rationality restricted to tests discussed in the feedback? This question is considered in chapter 3.
- Does feedback lead to a reduction of the number of test requests?
- Does feedback improve the rationality of test requests? This question and that above are answered in chapter 4.

- Is routine feedback capable of achieving a persistent change in test ordering behaviour? This question is answered in chapter 5.
- Do all general practitioners respond to the feedback to the same extent, and if not, what factors can explain why this response is not uniform?
- Beside the feedback, what was the influence of literature and continuing medical education on test ordering behaviour of general practitioners in the Maastricht region? This and the question above are dealt with in chapter 6.
- Does feedback (and the ensuing change in test ordering behaviour) have an adverse side effect on referral behaviour? This potential side effect is discussed in chapter 7.
- What are the economic consequences of feedback? They are shown in chapter 8.

References

1. Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *Br Med J* 1991; 303: 398-402.
2. Grossman RM. A review of physician cost-containment strategies for laboratory testing. *Med Care* 1983; 21: 783-801.
3. Horder J, Bosanquet N, Stocking B. Ways of influencing the behaviour of general practitioners. *J R Coll Gen Pract* 1986; 36: 517-21.
4. Sprij B, Casparie AF, Grol RPTM. Interventiemethoden om een verandering in de medische praktijkvoering te bewerkstelligen; wat is effectief? *Ned Tijdschr Geneesk* 1989; 133: 1115-8.
5. Buntinx F, Winkens RAG, Grol RPTM, Knottnerus JA. Influencing diagnostic and preventive performance in ambulatory care by feedback and reminders. A review. *Fam Pract* 1993; 10: 219-28.
6. Fraser CG, Woodford FP. Strategies to modify the test-requesting patterns of clinicians. *Ann Clin Biochem* 1987; 24: 223-31.
7. Axt-Adam P, van der Wouden JC, van der Does E. Influencing behavior of physicians ordering laboratory tests: a literature study. *Med Care* 1993; 31: 784-94.
8. Everett GD, de Blois CS, Chang PF, Holets T. Effect of cost education, cost audits, and faculty chart review on the use of laboratory services. *Arch Intern Med* 1983; 143: 942-4.
9. Fowkes FGR, Evans KT, Hartley G, Nolan DJ, Roberts CJ, Davies ER, et al. Multicentre trial of four strategies to reduce use of a radiological test. *Lancet* 1986; i: 367-70.

10. Gortmaker SL, Bickford AF, Mathewson HO, Dumbaugh K, Tirrell PC. A successful experiment to reduce unnecessary laboratory use in a community hospital. *Med Care* 1988; 26: 631-42.
11. Sommers LS, Sholtz R, Shepherd RM, Starkweather DB. Physician involvement in quality assurance. *Med Care* 1984; 22: 1115-38.
12. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus JA. Effect of feedback on test ordering behaviour of general practitioners. *Br Med J* 1992; 304: 1093-6.
13. Winickoff RN, Coltin KL, Morgan MM, Buxbaum RC, Barnett GO. Improving physician performance through peer comparison feedback. *Med Care* 1984; 22: 527-34.
14. Schectman JM, Elinsky EG, Pawlson LG. Effect of education and feedback on thyroid function testing strategies of primary care clinicians. *Arch Intern Med* 1991; 151: 2163-6.
15. Tierney WM, McDonald CJ, Martin DK, Hui SL, Rogers MP. Computerized display of past test results. Effect on outpatient testing. *Ann Intern Med* 1987; 107: 569-74.
16. Young DW. An aid to reducing unnecessary investigations. *Br Med J* 1980; 2: 1610-1.
17. Marton KI, Tul V, Sox HC. Modifying test-ordering behavior in the outpatient medical clinic. A controlled trial of two educational interventions. *Arch Intern Med* 1985; 145: 816-21.
18. Berwick DM, Coltin KL. Feedback reduces test use in a health maintenance organization. *JAMA* 1986; 255: 1450-4.
19. Martin AR, Wolf MA, Thibodeau LA, Dzau V, Braunwald E. A trial of two strategies to modify test-ordering behavior of medical residents. *N Engl J Med* 1980; 303: 1330-6.
20. Everett GD. Impact of supervision by medical teachers and in-patient test control programmes on the out-patient test utilization of residents. *Med Educ* 1985; 19: 138-42.
21. Tierney WM, McDonald CJ, Hui SL, Martin DK. Computer predictions of abnormal test result. Effects on outpatient testing. *JAMA* 1988; 259: 1194-8.
22. Eisenberg JM. An educational program to modify laboratory use by house staff. *J Med Educ* 1977; 52: 578-81.
23. Wilson GA, McDonald CJ, McCabe GP. The effect of immediate access to a computerized medical record on physician test ordering: a controlled clinical trial in the emergency room. *Am J Publ Health* 1982; 72: 698-702.
24. Wones RG. Failure of low-cost audits with feedback to reduce laboratory test utilization. *Med Care* 1987; 25: 78-82.

25. Long MJ, Cummings KM, Frisof KB. The role of perceived price in physicians' demand for diagnostic tests. *Med Care* 1983; 21: 243-50.
26. Schroeder SA, Myers LP, McPhee SJ, Showstack JA, Simborg DW, Chapman SA, et al. The failure of physician education as a cost containment strategy. *JAMA* 1984; 252: 225-30.
27. Cohen DI, Jones P, Littenberg B, Neuhauser D. Does cost information availability reduce physician test usage? *Med Care* 1982; 20: 286-92.
28. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med* 1990; 322: 1499-504.
29. Cummings KF, Frisof KB, Long MJ, Hryniewicz G. The effects of price information on physicians' test ordering behavior. *Med Care* 1982; 20: 293-301.
30. Applegate WB, Bennett MD, Chilton L, Skipper BJ, White RE. Impact of a cost-containment educational program on housestaff ambulatory clinic charges. *Med Care* 1983; 21: 486-96.
31. Shaw CD. Acceptability of audit. In: Smith R. *Audit in action*. London: BMJ Publishers, 1992.
32. Griner PF. Use of laboratory tests in a teaching hospital: long-term trends. Reductions in use and relative cost. *Ann Intern Med* 1979; 90: 243-8.
33. Nelson AR. Orphan data and the unclosed loop: a dilemma in PSRO and medical audit. *N Engl J Med* 1976; 295: 617-9.
34. Grivell AR, Forgie HJ, Fraser CG, Berry MN. Effect of feedback to clinical staff of information on clinical biochemistry requesting patterns. *Clin Chem* 1981; 27: 1717-20.
35. Schroeder SA, Kenders K, Cooper JK, Piemme TE. Use of laboratory tests and pharmaceuticals. Variation among physicians and effect of cost audit on subsequent use. *JAMA* 1973; 225: 969-73.
36. Eisenberg JM, Nicklin D. Use of diagnostic services by physicians in community practice. *Med Care* 1981; 19: 297-309.
37. Epstein AM, McNeil BJ. Physician characteristics and organizational factors influencing use of ambulatory tests. *Med Decis Making* 1985; 5: 401-15.
38. Ornstein SM, Markert GP, Johnson AH, Rust PF, Afrin LB. The effect of physician personality on laboratory test ordering for hypertensive patients. *Med Care* 1988; 26: 536-43.
39. Metcalfe DHH. Medical Audit. *Br Med J* 1974; iii: 327.
40. Kroenke K, Hanley JF, Copley JB, Matthews JI, Davis CE, Foulks CJ, et al. Improving house staff ordering of three common laboratory tests. Reduction in test ordering need not result in underutilization. *Med Care* 1987; 25: 928-35.
41. Eisenberg JM. Physician utilization. The state of research about physicians' practice patterns. *Med Care* 1985; 23: 461-83.

1.3 An outline of the study design

In this paragraph a brief overview of the design of the different study parts is given. A detailed description of the design of each study part is given in the undermentioned corresponding chapters.

The first part of the study was a retrospective analysis of the first results of the feedback, achieved in the period 1985-1989. Since internal controls were not available, Maastricht data were compared with data from a comparable reference laboratory elsewhere in the Netherlands. Data from the period 1983-1989 were comparable for 46 tests. Within this group of 46 tests several tests could be studied separately since they were discussed frequently in the feedback. The first effects of feedback were studied especially for these tests. Also it could be determined whether the influence of feedback was restricted only to tests discussed (see the first study question). Perhaps the feedback also influenced tests not discussed. To ensure anonymity, the data from the other laboratory were indexed to the number of patients in the Maastricht region.

The second part of the study was a randomized controlled trial. It aimed at determining a causal relation between feedback and a change in the volume of test requests as well as an improvement of the rationality of test ordering behaviour (the second and third study question). Over a 2½-year period, feedback was provided on tests not discussed before. These tests were split up into two test-groups which were, with regard to the provision of feedback, assigned to each of two random samples of general practitioners. One group of general practitioners received feedback on requests of endoscopy, cervical smears, electrocardiography and allergy tests. The other group received feedback on several radiological and ultrasonographical tests. The feedback on these two groups of tests was simply appended to the "usual" feedback. Mutually, for tests on which no feedback was provided, each group of general practitioners functioned as control group. Effects on the volume of requests and the rationality of requests were studied. Rationality of requests was determined by comparing each request and the accompanying clinical data with accepted guidelines. The methodology of the trial is discussed separately in chapter 9.

The third part of the study was an observational study on the long-term effects of feedback over the period 1983-1991, in which the persistence of the change in test ordering behaviour was assessed (the fourth study question).

The feedback that was provided prior to the randomized controlled trial was continued (and is still continued). Long-term trends of tests discussed frequently in the feedback were compared with trends of the remaining tests that were discussed not frequently or not at all. Within the group of tests that were frequently discussed special attention was given to those tests that were not simply discouraged but for which an alternative was suggested. The trend of these tests was compared with the trend of frequently discussed tests for which no alternative was suggested. For illustration of the long-term effects several examples of tests were studied. The extent of the change was determined by a new comparison of Maastricht data with data from a laboratory elsewhere in the Netherlands.

The fourth part of the study was a link between request data and a questionnaire to trace explanatory factors and confounders for the response to feedback. Response means the change (in terms of percentage) in the number of requests of tests frequently discussed in the feedback, before (1984) and after (1987) the start of the feedback. We tried to find factors that explained why doctors have different changes in test ordering behaviour through the (generally uniform) feedback (the fifth study question). Also we tried to discover other sources of information that, apart from the feedback, might have influenced test ordering behaviour (the sixth study question).

In an interview the questionnaire was filled in by the general practitioners who had received the feedback reports. The questionnaire focused on several practice and practitioner characteristics, such as diagnostic test facilities in the own practice, practice size, duration of consultations, practice setting, experience of the general practitioner, conception of his role, attitude to risk-taking, etcetera. Also it was studied to what extent the general practitioners had been reading literature or attended continuing medical education courses.

Finally, a survey was made of what the practitioner's opinion is of the Diagnostic Coordinating Centre Maastricht and its feedback. The data from the questionnaire were combined with request data per practitioner.

First, the data obtained by the questionnaire were related to test ordering behaviour, regardless of the effects of the feedback. Request data from a year prior to the interview were analyzed to that end. Second, the questionnaire data were related to the response to the feedback. The response was determined as described above. The general practitioners were ranked according to their response to the feedback. It was also studied to what extent the perception of the general practitioners about their own change in test ordering behaviour had any predictive value

for the actual change as assessed from the data of the Diagnostic Coordinating Centre Maastricht.

The fifth part of the study was an observational study on adverse side effects, meaning an increase in the number of referrals to compensate a lower use of diagnostic tests (the seventh study question). Referral data over the period 1985-1992 were obtained from the regional health insurance service VGZ. These referral data were compared with national referral data and with request data from the Diagnostic Coordinating Centre Maastricht. Especially the first referrals over the period 1986-1991 were studied. Possible adverse effects of feedback on referral trends were studied by looking at specialisms whose referrals would be susceptible for a change due to feedback (internal medicine, orthopaedic surgery). Also the referral rates from good responders to feedback were compared with referral rates from poor responders to feedback.

The sixth part of the study was an economic evaluation of the changes in request patterns brought about by the feedback. Is feedback, despite the extra costs of the feedback procedure itself, economically worthwhile (the eighth study question)? *The expenditure (charges) for diagnostic testing reflects only the financial impact of feedback, but for the economic consequences a different look at costs is important.*

In this study part, we split up costs per test into two major categories: fixed costs, being independent of the number of requests (such as manpower) and variable costs (chemicals, disposables, interest and depreciation of equipment). Especially the variable costs are considered.

Two calculations were made. In the first calculation, the trend in variable costs in the Maastricht region was compared with the same trend in the region of the reference laboratory, the latter reflecting the trend if no feedback had been provided. Trends were comparable for 44 tests.

A second calculation was based on an economic comparison of the total request trend in the Maastricht region with the national trend.

Based on these two calculations a model was made to predict the economic consequences for the implementation of feedback elsewhere.

A diagnostic centre for general practitioners: results of individual feedback on diagnostic actions

P. Pop and R.A.G. Winkens

Diagnostic Coordinating Centre Maastricht, the Netherlands

Background: In an earlier questionnaire study the general practitioners had indicated that they would appreciate individual comments after each group consultation (theoretical CHIME, thesis, University of Limburg, 1989). In 1992 the Diagnostic Coordinating Centre (DCC) was established as a general practitioners' centre for the purpose of diagnostic help, in order to create a more rational and efficient use of diagnostic facilities.

Objectives of the study: General practitioners can consult specialists at a Diagnostic Coordinating Centre (DCC) for the coordinated, the general practitioners' advice for detailed information which the specialists have to provide for the general practitioner.

Methods: The following different forms of programmes were used: individual and group consultations, individual and group consultations, and the relation between primary and secondary health care.

Results of general practitioners

In 1992 a part of the 50 general practitioners in the region received a letter concerning critical comments on the appropriateness of their requests for diagnostic tests during a month chosen at random. The comments were given by the DCC. The results of the diagnostic centre were as follows in the first month.

The DCC was created in 1992 as a result of the application from which are completed by general practitioners when requesting a diagnostic test. The application form gives the description of the clinical situation, history, symptoms, signs, and

Summary

A diagnostic centre, managing diagnostic tests for general practice, can improve the service provided by primary health care and the communication between general practitioners and specialists. In addition, it can evaluate the use and misuse of tests. This paper describes the work of a diagnostic centre in the Netherlands serving 80 general practitioners. Following the introduction of individual feedback to general practitioners on their use of diagnostic tests there was a decrease in the number of requests for tests.

The first part of the study was an economic evaluation of the change in request patterns brought about by the feedback. It showed the extra costs of the feedback procedure itself, compared with the savings in the use of tests. The second part of the study was a descriptive study of the use of tests. It showed that the number of requests for tests had decreased after the introduction of feedback. The third part of the study was a descriptive study of the use of tests. It showed that the number of requests for tests had decreased after the introduction of feedback.

In this study part, we light up the use of tests and the costs of the tests. The costs of the tests are divided into fixed costs and variable costs. The fixed costs are the costs of the equipment and the variable costs are the costs of the tests themselves. The results show that the variable costs are much higher than the fixed costs.

The results of the study show that the number of requests for tests has decreased after the introduction of feedback. This is a positive result, as it means that the diagnostic centre is being used more efficiently. The results also show that the costs of the tests are much higher than the fixed costs, which means that the variable costs are a significant part of the total costs.

A second evaluation was carried out to see if the results of the first evaluation could be confirmed. The results of the second evaluation were similar to the results of the first evaluation, which confirms the findings of the first evaluation.

Based on these two evaluations, we can conclude that the introduction of feedback to general practitioners has led to a decrease in the number of requests for tests. This is a positive result, as it means that the diagnostic centre is being used more efficiently.

Introduction

There has been a diagnostic centre in Maastricht, the capital of the Dutch province of Limburg, since 1979. The aims of the centre are to improve the service provided by primary health care and the cooperation between general practitioners and specialists. The centre is located in the university hospital of Maastricht, the only hospital in the area, and serves 80 general practitioners, responsible for a population of approximately 186000. The diagnostic centre has a wide range of activities:

- Composing protocols for diagnostic procedures.
- Evaluating the diagnostic work-up* of each general practitioner and providing individual feedback. In an earlier questionnaire study the general practitioners had indicated that they would appreciate individual comments rather than group comments (Beusmans GHMI, thesis, University of Limburg, 1986). In 1985 the diagnostic centre started to provide feedback to every general practitioner about his or her use of diagnostic tests, in order to create a more rational and efficient use of diagnostic facilities.
- Consultations by telephone. General practitioners can consult specialists at a time that is convenient for both. Prior to the consultation, the general practitioner is asked for detailed information which the specialist uses to prepare himself for the consultation.
- Initiating and supporting different forms of postgraduate education.
- Initiating and carrying out scientific studies focussing on the relation between primary and secondary health care.

Feedback to general practitioners

At least twice a year each of the 80 general practitioners in the region receives a report containing critical comments on the appropriateness of their requests for diagnostic tests during a month chosen at random. The comments are given by the coordinator of the diagnostic centre who is a specialist in internal medicine.

The critical analysis is based on the application forms which are completed by general practitioners when requesting a diagnostic test. The application form lists the categories of tests -clinical chemistry, haematology, serology, virology, faeces tests, urine tests, bacteriology, electrocardiography, radiology, ultrasonography, endoscopy and cytology/histology- and gives details of tests that are available. It

also asks for information about the patient including signs, symptoms, possible diagnosis, previous history, drugs prescribed and the reason(s) for the request- to exclude diagnosis, to confirm diagnosis, screening, check-up, requested by patient, checking a known disorder, for reassurance.

The report normally includes the following:

- @ The total number of tests requested by the general practitioner during the month and a comparison with the number of requests in a previous analysis and with the mean number of requests from all general practitioners.
- @ A review of the quality and quantity of information about the patient supplied on the application form.
- @ A discussion about diagnostic tests in general and about several individual patients (their names and date of birth are provided, giving the general practitioner the opportunity to look at the patients' files).
- @ Comments on strikingly frequent or infrequent requests; incorrect or redundant requests in relation to certain signs or diseases or for specific groups of patients; defined combinations of tests for renal function, liver function, thyroid function and so on; tests which are unnecessary because they result in no change in the general practitioner's management; and the cost of redundant tests.

In addition, test requests for several patients are discussed with special regard to the relevance of requested tests for the complaints or physical signs mentioned; advice about appropriate tests; and suggestions for a different diagnostic work-up. The report also asks questions about the general practitioner's management after receiving normal or abnormal test results.

For example:

"Based on the history (cough, haemoptysis, weight loss, heavy smoking) and results of a physical examination you strongly suspected that Mr A, born in 1934, has lung cancer. For this reason you requested a chest X-ray. The test result was negative. In the light of your suspicion, what was your management after this test result?"

The general practitioners are encouraged to return their answers and to comment on the remarks and suggestions made in the report.

Descriptive study

The suggestions for improvement and remarks contained in the reports are mainly concerned with tests in the fields of clinical chemistry, haematology, serology and bacteriology and with faeces and urine tests. Some general practitioners receive the same comments repeatedly because they continue with a fixed recognizable pattern of requests. Therefore, to discover if there had been any chronological change in patterns of diagnostic requests in the above categories, a descriptive study was performed and during the period 1979-1987, data from the diagnostic centre were analyzed.

Results of feedback

Most general practitioners appreciated the reports provided by the diagnostic centre. On average 70% answered the questions or gave comments. The reports appeared to have had a marked influence on diagnostic performance. From 1979 to 1984 the number of requests increased but in 1985 there was a slight decrease and in 1986 and 1987 this decrease was much larger (table 1).

If the diagnostic tests are examined more closely the influence of the feedback can be seen (table 2). For example, when determining renal function, tests for serum creatinine level and serum urea level were usually used. It was repeatedly stated in the feedback that the serum creatinine level was sufficient and as a result of this the number of serum urea determinations for the 80 doctors decreased from 2883 in 1984 to 256 in 1987 (table 2).^{1,2}

For complaints about joints serological tests for rheumatic diseases were often requested -mostly at the instance of the patient. In the feedback it was repeatedly stated that Rose-Waaler and latex fixation tests have a low predictive value for excluding or demonstrating rheumatoid arthritis in general practice. Over the period 1984-1987 the number of Rose-Waaler and latex fixation tests decreased from 1188 and 1133 respectively to 406 and 381 respectively (table 2).^{3,4}

Patients less than 40 years of age with vague and non-specific complaints and without abnormal findings on physical examination, regularly underwent several screening procedures. A frequent combination was: erythrocyte sedimentation rate, blood smear, serum glucose, serum creatinine and liver function tests. In this group of patients, abnormal results were rare and general practitioners were advised to omit this group of screening tests as far as possible. Between 1984 and 1987 the number of requests for these tests decreased (table 2).

24 Table 1. Total number of applications for diagnostic tests over the period 1979-1987 by the 80 general practitioners^a

Type of test	Number of applications							
	1979 ^b	1980 ^c	1982 ^d	1983 ^b	1984 ^b	1985 ^b	1986 ^e	1987 ^b
Clinical chemistry	46156	44328	49108	45537	51775	46569	39576	34928
Haematology	44172	41236	48072	52948	59556	53437	45913	38496
Serology/Virology	10329	8572	8724	8192	8554	7329	5518	4866
Urine/faeces tests	2212	1848	2524	3858	4137	3559	3328	2918
Bacteriology	1256	1272	2172	1679	1845	2558	2508	2614
Total	104125	97266	110600	112214	125867	113452	96843	83822

^aData for 1981 were not analysed. ^bRecorded over 12 months. ^cRecorded over three months; ^dfour months; ^eeight months and adjusted to 12 months.

Table 2. Number of requests by the 80 doctors for tests discussed in feedback in 1984 and 1987

Test	Number of request		Percentage decrease
	1984	1987	
Serum urea	2883	256	91
Rose-Waaler test	1188	406	66
Latex fixation test	1133	381	66
Blood smear	11320	5920	48
Alkaline phosphatase level	5347	3010	44
Aspartate aminotransferase level	2207	1232	44
Alanine aminotransferase level	5545	3187	43
White blood-cell count	11753	7007	40
Gamma glutamyl transpeptidase level	6136	4036	34
Haemoglobin	12293	8373	32
Erythrocyte sedimentation rate	13749	10096	27
Serum creatinine	5370	4208	22
Serum glucose	8224	7450	9

Discussion

The results of this study show an overall decrease in the number of requests for diagnostic tests between 1979 and 1987. This was not the main aim of the diagnostic centre but a more rational and efficient use of diagnostic tests does seem to lead to lower use. In the feedback provided to the doctors, comments are given on obvious instances of unnecessary requests and the fall in these unnecessary requests can indeed be seen as a positive effect. Not all general practitioners respond to feedback in the same way. A questionnaire survey revealed a strong positive correlation between a positive appreciation of protocols from the diagnostic centre and a low use of diagnostic tests.⁵ The findings of this study, however, reveal only limited information about the effects of feedback and further research is needed.

The number of requests is influenced by many factors. For example, the decrease in the number of serum glucose tests was less than expected, presumably because in recent years there had been a move to look after patients with diabetes mellitus type 2 in primary rather than secondary care.

To what extent individual feedback caused the decrease in use of diagnostic tests seen in this study and to what extent other factors, such as postgraduate education contributed is not yet clear. Research in this field is needed and a study, in cooperation with the Ministry of Public Health is in preparation.

References

1. Pop P, Winkens RAG, van Leer JVM. Determination of renal function (1). Practitioner (Dutch edition) 1987; 4: 449-52.
2. Pop P, Winkens RAG, van leer JVM. Determination of renal function (2). Practitioner (Dutch edition) 1987; 4: 536-8.
3. Pop P, van Leer JVM, Fijten G. Serologic tests for the diagnosis of rheumatoid arthritis (1). Practitioner (Dutch edition) 1987; 4: 904-7.
4. Pop P, van Leer JVM, Fijten G. Serologic tests for the diagnosis of rheumatoid arthritis (2). Practitioner (Dutch edition) 1987; 4: 997-9.
5. Beusmans GHMI, Does RJMM. Working arrangements of the diagnostic centre Maastricht. Huisarts Wet 1988; 31: 154-8.

Acknowledgement

The authors thank J. Horder for his advice and critical review of the manuscript.

work-up: any or all of the procedures used to assemble data and present a collective picture of a patient's condition in order to reach a diagnosis, often a specific diagnosis to determine a particular course of treatment. The taking of a medical history and the administration of a physical examination and laboratory tests are standard features, but often other specialized examinations are conducted as well. Source: International dictionary of medicine and biology. New York: John Wiley, 1986.

Effect of feedback on test ordering behaviour of general practitioners

R.A.G. Winkens¹, P. Pop¹, R.P.T.M. Grol², A.D.M. Kester³ and J.A. Knottnerus²

- ¹ Diagnostic Coordinating Centre Maastricht
- ² Department of General Practice, University of Limburg, Maastricht
- ³ Department of Methodology and Statistics, University of Limburg, Maastricht

Abstract

Objective

To assess the effects of feedback on the test ordering behaviour of general practitioners.

Design

Comparison of requests at two diagnostic centres, and internal comparison between tests which were discussed in feedback and tests which were not.

Setting

A diagnostic centre in Maastricht giving feedback and another elsewhere in the Netherlands (laboratory A) not giving feedback.

Subjects

All 85 general practitioners in the region of Maastricht, and all general practitioners in the region of laboratory A.

Main outcome measures

Numbers of tests requested by general practitioners.

Results

Requests at the Maastricht diagnostic centre decreased soon after the onset of feedback whereas there was a persistent increase in requests at laboratory A. Tests that were discussed showed the strongest decrease (maximum 40%), though tests that were not discussed decreased as well (maximum 27%).

Conclusions

Feedback on diagnostic requests may exert a strong influence on test ordering behaviour. Four years after the onset of feedback the effects were still noticeable.

Introduction

For many years there have been efforts in primary and secondary health care to influence the behaviour of physicians.¹⁻⁵ One method is to provide feedback⁶, and the effects have been extensively studied.⁷⁻¹⁶ Results are generally positive, although the effects are usually temporary.^{8,11,13-16} In reported series, however, feedback was offered mainly within the setting of scientific studies with limited duration.

The diagnostic centre in Maastricht provides feedback as a health care activity.¹⁷ Uniquely based on the request forms filled in by the general practitioner, comments are made about inappropriate requests or recommendations are offered with regard to rational diagnosis. Thus the feedback concerns real cases from daily general practice.

Overviews and estimations of the Maastricht diagnostic centre data indicated that the feedback was accompanied by a reduction in the total number of tests ordered in subsequent years. In these first analyses no account was taken of factors such as the intensity of the information provided in the feedback (the more frequently a subject is discussed the greater is the effect expected) and the request trends which would have occurred in the absence of feedback. These considerations prompted further investigation, in which we addressed the following questions: (1) What trends may be discerned in the test ordering behaviour of general practitioners over time at the Maastricht diagnostic centre? (2) Are the request patterns at the centre different from those of laboratories where no feedback is provided? (3) Do the changes observed in test ordering behaviour reflect advice given in the feedback? (4) What are the financial implications of any differences in request patterns resulting from the feedback?

Subjects and methods

Since 1985 general practitioners served by the diagnostic centre in Maastricht have been provided with feedback as critical biannual reports. These reports, which are based on analyses of request forms submitted in one month, discuss both quantitative and qualitative aspects of diagnostic test ordering. As the request forms offer clinical data on the patient (history, physical findings, suspected diagnosis, etc.) it is feasible to assess the rationality of tests ordered.

The following data were gathered. Firstly, we surveyed the requests during 1983 to 1990, then made an inventory of tests which were discussed in the feedback

reports from 1985 to 1990. We also obtained data on test requests during 1983 to 1989 in another Dutch region where a laboratory processed requests for diagnostic tests from general practitioners. That reference laboratory (referred to here as laboratory A) and its region were comparable to the Maastricht diagnostic centre and its catchment area with regard to the degree of urbanisation and the average family practice size. The Maastricht diagnostic centre and laboratory A processed virtually all requests for tests from general practitioners in the two regions. General practitioners in the region served by laboratory A received no feedback of the kind described in this paper.

Of all tests that were available, there were 46 that could be requested from both the diagnostic centre in Maastricht and laboratory A. The two centres were compared for these 46 tests, which are listed in the table. Eleven tests were discussed in virtually all reports to general practitioners, often more than once (see table). Hence any influence of feedback on test ordering behaviour should have been particularly evident for these tests.

The data were analyzed with respect to the total number of requests for the 46 tests available in both centres, the number of requests for tests discussed frequently in feedback reports (see table) and the number for tests that were not, and several typical examples from the group of tests discussed frequently. To assess the economic implications we also carried out a financial appraisal, in which costs with and without feedback were compared.

To ensure anonymity of laboratory A all the data presented were indexed to the level of the Maastricht diagnostic centre in 1984. Beginning with 1985 we used the X^2 test to examine the data for each year to see whether differences were significant.¹⁸ The request figures for each year were compared with those for 1984. Figures for 1984 counted as zero measurement.

Results

Figure 1 illustrates the request trends at the Maastricht diagnostic centre with respect to both the total number of requests yearly and requests for the 46 tests also available in laboratory A.

Figure 2 shows the yearly numbers of requests for the 46 tests alone at the two centres and in laboratory A with the data indexed to the number of patients in the Maastricht diagnostic centre's region (curve x).

Tests available in both Maastricht diagnostic centre and laboratory A

Clinical chemistry		Haematology	Serology	Faecal tests
Blood	Urine			
Glucose	Glucose, qual.	Haemoglobin*	Pregnancy test.	Occult blood
Sodium	Urobilinogen, qual.	Packed cell volume*	Monospot test.	Absorption
Potassium	Bilirubin, qual.	Erythrocyte sedimentation rate*	Rubella	Worm egg
Chloride	Acetone, qual.	Leucocyte count*	haemagglutination	Cysts
Calcium	Sediment	Differential count*		
Urea*	Glucose, quant.	Erythrocyte count*		
Creatinine	Amylase, qual.	Platelet count		
Urate	Creatinine, quant.			
Alkaline phosphatase	Vanillylmandelic acid, quant.			
Acid phosphatase	Albumin, qual.			
γ -Glutamyltransferase				
Bilirubin				
Aspartate aminotransferase*				
Alanine aminotransferase*				
Lactate dehydrogenase*				
Creatine phosphokinase				
Amylase				
Total protein				
Protein electrophoresis				
Iron*				
Triglyceride				
Cholesterol				

* Tests (n=11) marked with asterisk were those discussed in virtually all reports to general practitioners, often more than once.

Though the number of requests at the two centres were commensurate with the size of the patient population, the number of requests in the Maastricht diagnostic centre was considerably lower even before feedback was started (figure 2). The reduction in the number of requests for the 46 tests at the Maastricht diagnostic centre was clearly noticeable from 1985 onwards (curve z). By contrast, the number of tests ordered from laboratory A increased continually. The differences were significant ($p < 0.001$). Even when numbers for the two laboratories were standardised for 1984 (curve y versus curve z) differences remained significant ($p < 0.001$).

Figure 1. Total numbers of tests requested yearly at Maastricht diagnostic centre and numbers of requests for 46 tests also available in lab. A

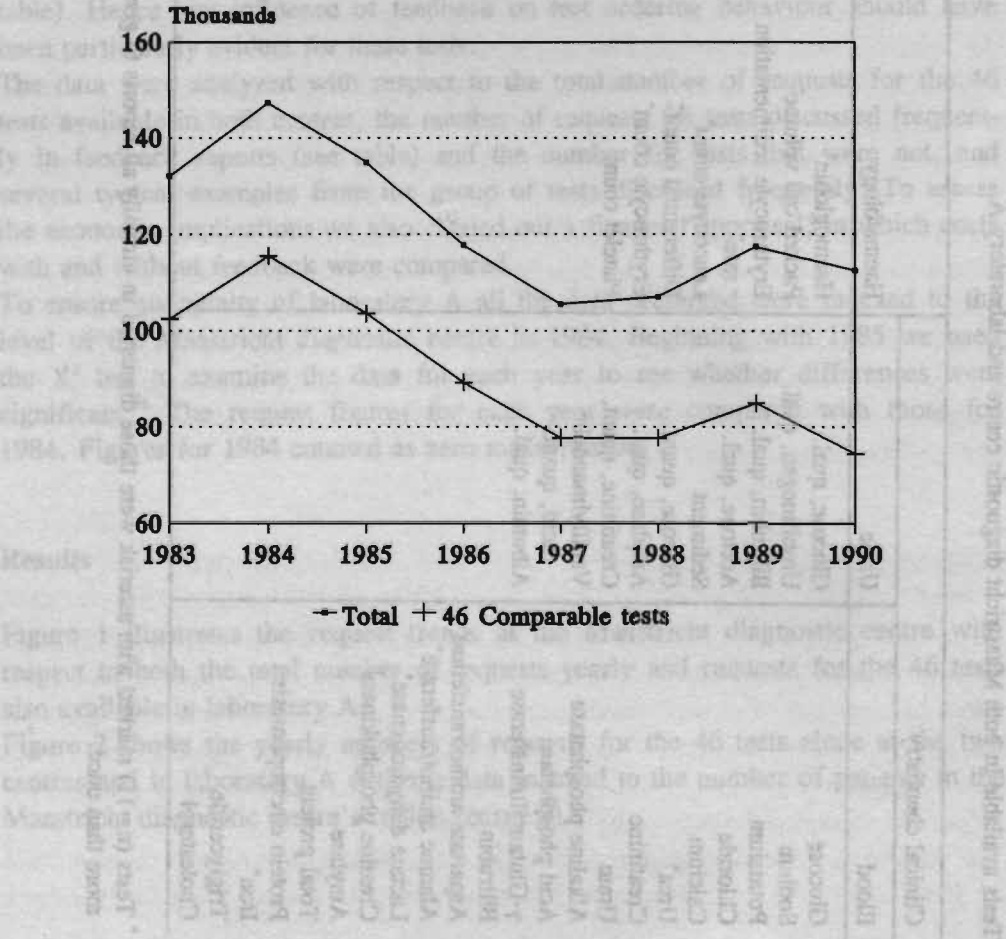


Figure 2. Yearly numbers of requests for 46 tests available in Maastricht diagnostic centre and laboratory A and in laboratory A with data indexed to numbers of patients in Maastricht diagnostic centre region (curves x, y, z; see text)

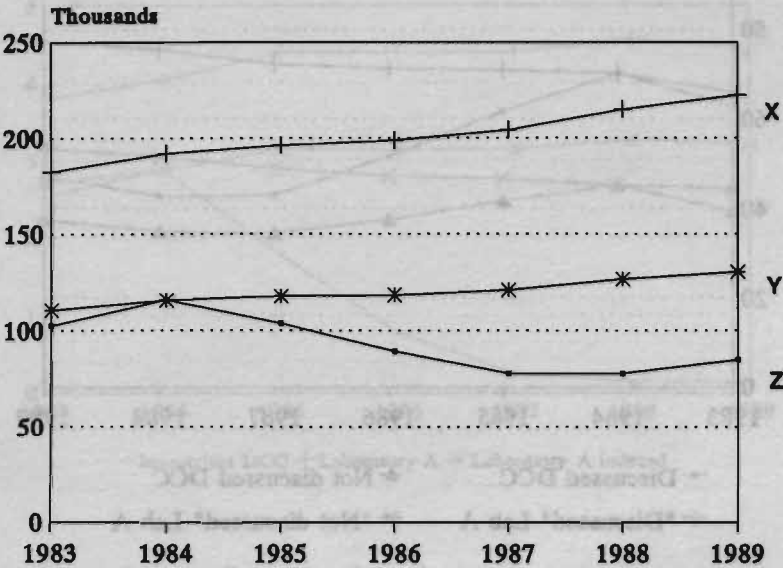
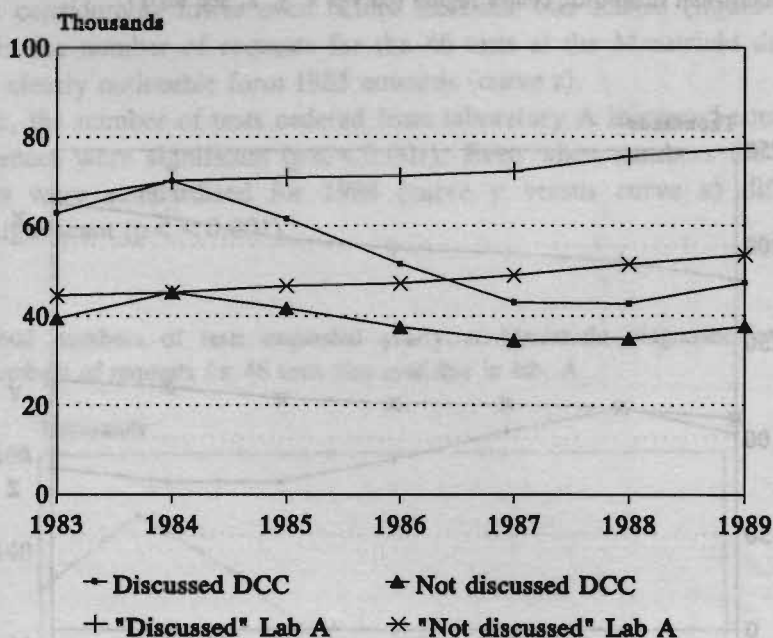


Figure 3 shows the 46 tests available in the two centres classified by whether they were (n=11) or were not (n=35) frequently discussed in the feedback. Requests for the 11 tests frequently discussed are thus compared for the two laboratories. The effect of feedback was striking for these tests (for every year $p < 0.001$), but requests for the 35 tests not discussed frequently also decreased significantly in Maastricht ($p < 0.001$). On examining data for Maastricht alone with respect to request trends for tests frequently discussed in the feedback and tests that were not dealt with in the reports, the decrease was seen to be most prominent for tests that were discussed (figure 3; $p < 0.001$).

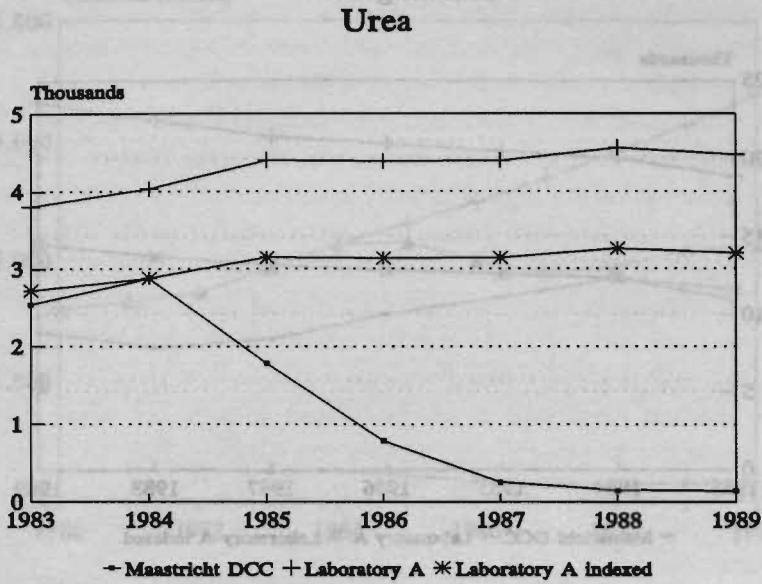
To illustrate the effect more clearly we elaborated a few examples that we expected would reveal a large impact of feedback (figure 4). These were tests stated repeatedly to be inappropriate under the given circumstances- namely, urea, lactate dehydrogenase, and haemoglobin measurements. These were compared with creatinine measurement, a test not designated as inappropriate.

Figure 3. Forty six tests available in two centres classified by whether they were (n=11) or were not (n=35) discussed in feedback



The differences between the two laboratories were unequivocal ($p < 0.001$). The observed changes had major financial implications. For all diagnostic procedures (endoscopy not included) expenditure of the Maastricht diagnostic centre increased from about fl 1.59m (£500000) in 1980 to fl 2.14m (£673000) in 1984. Nationally, the annual increase was 8%. On the basis of this rate of increase the expenditure would have amounted to fl 3.39m (£1.07m) in 1990. In fact, costs were fl 2.13m (£670000). Included were the extra costs of the data processing and analyses for the feedback reports (figure 5).

Figure 4. Four examples of influence of feedback in Maastricht diagnostic centre and laboratory A



Lactate dehydrogenase

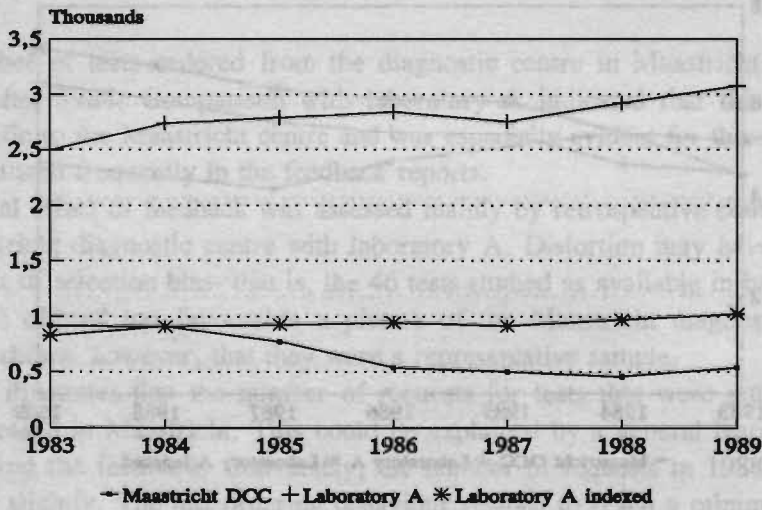
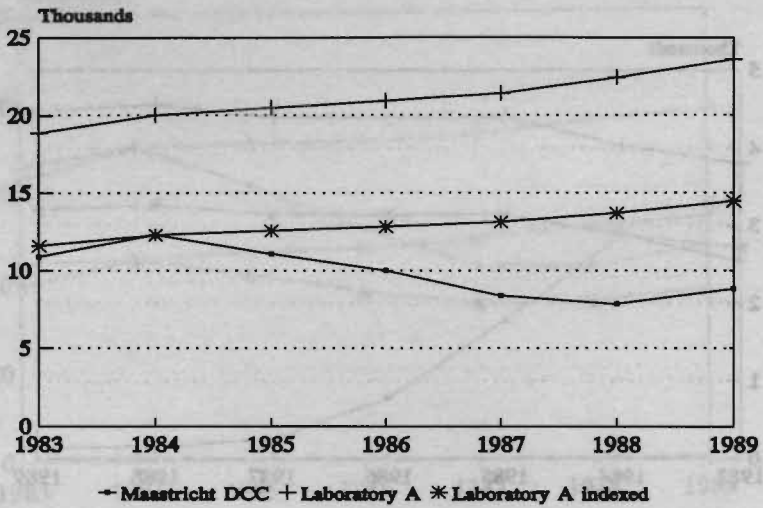


Figure 4 continued

Haemoglobin



Creatinine

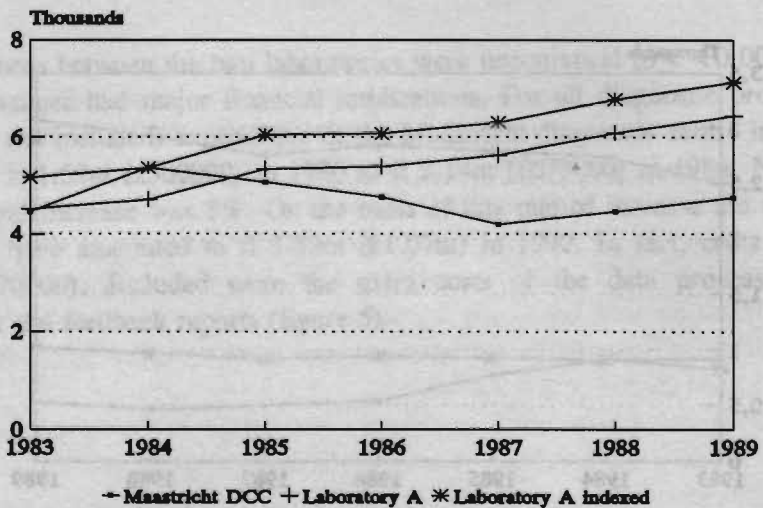
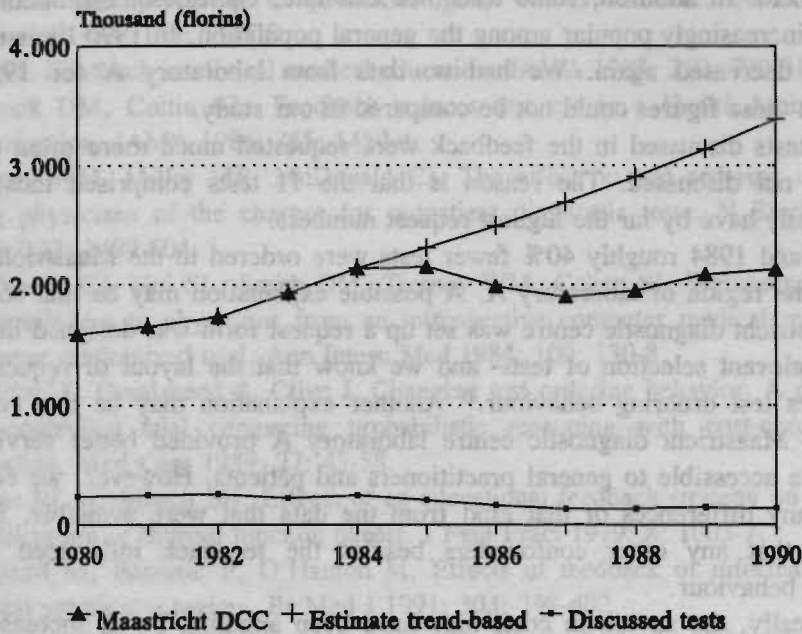


Figure 5. Annual costs of Maastricht diagnostic centre



Discussion

The number of tests ordered from the diagnostic centre in Maastricht decreased sharply after 1984. Comparison with laboratory A indicated that this reduction was specific to the Maastricht centre and was especially evident for those tests that were discussed frequently in the feedback reports.

A potential effect of feedback was assessed mainly by retrospective comparison of the Maastricht diagnostic centre with laboratory A. Distortion may have occurred as a result of selection bias- that is, the 46 tests studied as available in both centres may have offered too favourable a picture of the Maastricht diagnostic centre. Figure 1 shows, however, that they were a representative sample.

Figure 3 illustrates that the number of requests for tests that were not discussed also decreased in Maastricht. This could be explained by a general learning effect arising from the feedback. Conversely, the number of requests in 1988 and 1989 increased slightly. The test ordering behaviour seemed to reach a minimal level in 1987. Probably an increase in the number of requests might have been expected, given the reinforcement of primary health care in recent years- for instance, with

regard to the management of non-insulin dependent diabetes mellitus by general practitioners. In addition, some tests- for example, cholesterol estimations- have become increasingly popular among the general population. In 1990 the number of requests decreased again. We had no data from laboratory A for 1990, and therefore these figures could not be compared in our study.

The 11 tests discussed in the feedback were requested much more often than the 35 tests not discussed. The reason is that the 11 tests comprised those which individually have by far the highest request numbers.

In 1983 and 1984 roughly 40% fewer tests were ordered in the Maastricht region than in the region of laboratory A. A possible explanation may be that soon after the Maastricht diagnostic centre was set up a request form was designed that listed only a relevant selection of tests- and we know that the layout of request forms influences test ordering behaviour.¹⁹ Another explanation may be that compared with the Maastricht diagnostic centre laboratory A provided better services and was more accessible to general practitioners and patients. However, we could not deduce any differences of that kind from the data that were available. It seems unlikely that any other confounders besides the feedback influenced the test ordering behaviour.

Theoretically, the saving in costs may have been annulled by an increase in the number of referrals. Overall referral patterns did not indicate that this had occurred. Finally, although p values indicated highly significant differences, this was partly due to the large number of requests.

This study was granted by the Dutch Ministry of Public Health. We thank the members of its guiding committee for their assistance during the study.

References

1. Fraser CG, Woodford FP. Strategies to modify the test-requesting patterns of clinicians. *Ann Clin Biochem* 1987; 24: 223-31.
2. Horder J, Bosanquet N, Stocking B. Ways of influencing the behaviour of general practitioners. *J R Coll Gen Pract* 1986; 36: 517-21.
3. Grossman RM. A review of physician cost-containment strategies for laboratory testing. *Med Care* 1983; 21: 783-801.
4. Eisenberg JM, Williams SV. Cost Containment and changing physicians' practice behavior. *JAMA* 1981; 246: 2195-201.

5. Sprij B, Casparie AF, Grol R. Interventiemethoden om een verandering in de medische praktijkvoering te bewerkstelligen; wat is effectief? *Ned Tijdschr Geneesk* 1989; 133: 1115-8.
6. Ende J. Feedback in clinical medical education. *JAMA* 1983; 250: 770-81.
7. Berwick DM, Coltin KL. Feedback reduces test use in a Health Maintenance Organization. *JAMA* 1986; 255: 1450-4.
8. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med* 1990; 322: 1499-504.
9. McDonald CJ, Hui SL, Smith DM, Tierney WM, Cohen SJ, Weinberger M, et al. Reminders to physicians from an introspective computer medical record. A two-year randomized trial. *Ann Intern Med* 1984; 100: 130-8.
10. Davidoff F, Goodspeed R, Clive J. Changing test ordering behavior. A randomized controlled trial comparing probabilistic reasoning with cost-containment education. *Med Care* 1989; 27: 45-58.
11. Rhyne RL, Gehlbach SH. Effects of an educational feedback strategy on physician utilization of thyroid function panels. *J Fam Pract* 1979; 8: 1003-7.
12. Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *Br Med J* 1991; 303: 398-402.
13. Eisenberg JM. An educational program to modify laboratory use by house staff. *J Med Educ* 1977; 52: 578-81.
14. Fowkes FGR, Evans KT, Hartley G, Nolan DJ, Roberts CJ, Davies ER, et al. Multicentre trial of four strategies to reduce use of a radiological test. *Lancet* 1986; i: 367-70.
15. Marton KI, Tul V, Sox HC. Modifying test-ordering behavior in the out-patient medical clinic. A controlled trial of two educational interventions. *Arch Intern Med* 1985; 145: 816-21.
16. Bareford D, Hayling A. Inappropriate use of laboratory services: long term combined approach to modify request patterns. *Br Med J* 1990; 301: 1305-7.
17. Pop P, Winkens RAG. A diagnostic centre for general practitioners: results of individual feedback on diagnostic actions. *J R Coll Gen Pract* 1989; 39: 507-8.
18. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley and Sons, 1981.
19. Zaat JOM, van Eijk JThM, Bonte HA. Mag het ook een testje minder? De invloed van een beperking van het aanvraagformulier voor laboratorium-onderzoek. *Huisarts Wet* 1991; 34: 72-7.

This study was supported by a grant from the Dutch Ministry of Welfare, Health and Cultural Affairs (project 88-28).

Submitted for publication

A randomized controlled trial of routine individual feedback on diagnostic test requests

- an evaluation of quantity and quality changes -

R.A.G. Winkens¹, P. Pop¹, A.M.A. Bugter-Maessen¹, R.P.T.M. Grol², A.D.M. Kester³, G.H.M.I. Beusmans⁴ and J.A. Knottnerus⁴

¹ Diagnostic Coordinating Center Maastricht

² Center for Research on Quality in Family Practice, Universities of Nijmegen and Maastricht

³ Department of Methodology and Statistics, University of Limburg, Maastricht

⁴ Department of Family Practice, University of Limburg, Maastricht

The study was supported by a grant from the Dutch Ministry of Welfare, Health and Cultural Affairs (project 88-28).

Submitted for publication

Background

Feedback is frequently used to change test ordering behavior. Effects can be reductions in test usage and savings in costs. Since 1985, the Diagnostic Coordinating Center Maastricht has been giving feedback on diagnostic tests as a routine health care activity twice a year to all family physicians in its region. Both quantity and quality are discussed in this feedback. The effects of this feedback were studied.

Methods

In a randomized controlled trial over 2½ years feedback was given on tests not discussed in the feedback before. One (n=39) of two groups of family physicians received feedback on electrocardiography, endoscopy, cervical smears and allergy tests (test-group A). The other group (n=40) received feedback on radiologic and ultrasonographic tests (test-group B). Changes in volume and rationality of requests were evaluated.

Results

For test-group A the number of requests decreased for the intervention group (-7%) while this number increased for the control group (+13%) ($p=0.04$). On test level significant differences in favor of the intervention group were seen for cervical smears ($p=0.01$) and lumbar spine X-rays ($p=0.02$). Rationality improved for test-group B; non-rational requests were seen less frequently in the intervention group compared to the control group ($p=0.04$). On test level significant differences in rationality in favor of the intervention group were seen for lumbar spine X-rays ($p=0.004$).

Conclusions

Routine feedback can change the quantity and quality of requests.

Key words

Feedback, Diagnostic Tests, Quality Assurance, Family Practice, Randomized Controlled Trial.

Introduction

One of the methods for influencing the test ordering behavior of physicians is feedback. The effects of feedback have been extensively studied with contradictory findings.¹⁻¹⁰ Most of these studies were of limited duration. In one study the effects were followed after the feedback was stopped. The effects diminished soon thereafter.⁵ Frequently, the effects were expressed in terms of a reduction of the numbers of tests ordered or a saving in expenses for diagnostic testing.⁴⁻¹⁰ Apart from these quantitative aspects there is also a perhaps even more important qualitative aspect, namely the influence on the rationality or appropriateness of diagnostic requests. Until now, only a few studies have evaluated the effects of feedback on the quality of test ordering behavior.¹¹ Schectman et al. focused specifically on the compliance of test ordering behavior with guidelines for the use of thyroid function testing. Given the limited evidence so far, there is need for more research on the effects of feedback on the rationality of test ordering behavior. Also the coherence between changes in volume and rationality of diagnostic testing should be considered in research.

At the Diagnostic Coordinating Center Maastricht feedback on test ordering is provided as a routine health care activity. The main goal of this feedback is to improve the quality of test ordering behavior. Also, the feedback is a way to reduce the number of unnecessary requests. To that end, at set times a large number of requests from individual family physicians are discussed regarding their rationality and volume. Surveys showed that the introduction of the feedback was followed by a marked reduction of the number of requests in subsequent years. After two years, the total number of requests was reduced by 24%. A reduction, for some tests up to 95%, was particularly seen for tests discussed in the feedback and was specific for the region of the Diagnostic Coordinating Center Maastricht.¹² A causal relation between our feedback and a change in the volume and the rationality of requests was not yet demonstrated.

So far the effects of such a profound and personal type of feedback, provided routinely for many years (and not as part of a specifically organized, and often temporary, study setting), have not been studied in a randomized controlled trial. In such a trial the following questions were addressed:

1. Does routine individual feedback cause a reduction in the number of tests?
2. Does routine individual feedback cause an increase in the rationality of tests?

Methods

Background

Since 1985, feedback on diagnostic actions has been given twice a year to the (85) family physicians affiliated to the Diagnostic Coordinating Center Maastricht. These family physicians provide care for about 187000 patients. Every individual family physician receives a feedback report with critical comments on his or her requests. The feedback is provided by a respected expert (internist) and is based on an analysis of request forms filled out by the family physician. As a result the feedback concerns real cases from daily practice, discussing a large variety of diagnostic tests. The feedback reports discuss both the volume and the rationality of test requests submitted in the previous month. Since the request forms contain clinical data on the patient (history, physical signs and findings, suspected diagnosis etc.) it is feasible to assess the rationality of the tests ordered. Rationality is determined on the basis of accepted regional guidelines and standards of the Dutch College of General Practitioners, the NHG. Comments about inappropriate requests are made, and recommendations and alternatives are offered for more rational diagnostic testing. To assist the physician, patients are mentioned by name and date of birth. The details and contents of the feedback have been published elsewhere.¹³

Design of the trial

A randomized controlled trial was carried out over 2½ years (October 1989 until May 1992) among all family physicians affiliated to the Diagnostic Coordinating Center Maastricht. In this period, routine feedback was provided 5 times (in October 1989, April and November 1990, May and December 1991), discussing the requests from September 1989, March and October 1990, April and November 1991 respectively. Experimental feedback was added to the routine feedback. Experimental feedback was given on tests which were not, or only incidentally, discussed in the feedback before. Before the trial, feedback was mainly provided on a variety of blood, urine and feces tests. Other tests like radiology and electrocardiography were hardly discussed (in less than 5% of all feedback comments) and histology was not discussed.

The selected tests were split up into two test-groups. The first test-group (A) consisted of cervical smear, electrocardiography, endoscopy and allergy tests (IgE,

RAST, Phadiatop[®]). The second test-group (B) consisted of several radiologic and ultrasonographic tests (X-ray of the chest, cervical spine, thoracic spine, lumbar spine, pelvis, knees, ankles and sinuses and ultrasonography of the kidneys and liver/biliary tract). The selected tests could not be split up at random since they were not always independent from each other. For example, X-rays of the lumbar spine are often requested in combination with X-rays of other parts of the vertebral column or the pelvis. Moreover, feedback on tests in one group should not affect requests of tests in the other group.

Not all family physicians in our region could be included in the trial. Due to retirement from practice 6 of them were not working in practice during at least 6 months of the trial. All remaining family physicians (n=79) were divided at random into two groups. Group 1 received feedback on tests belonging to test-group A; group 2 received feedback on tests belonging to test-group B. Both groups functioned as each other's controls for tests which were discussed in the feedback provided to the other group. After randomization, groups appeared to be similar on factors like practice setting, degree of urbanization, and experience and gender of the family physicians (table 1).

Table 1. Practice and physician characteristics at randomization
Group 1 versus group 2

	Group 1	Group 2
Number of physicians	39	40
Doctor characteristics:		
- mean experience (in years)	15	15
- gender:		
* male	35	36
* female	4	4
Practice characteristics:		
- practice setting:		
* solo	21	23
* group practice	11	10
* health center	7	7
- urbanization:		
* city	28	25
* rural area	11	15

The family physicians had to be unaware of the trial. This was possible because the routine feedback as given since 1985 was continued. The new feedback on the tests discussed as part of the trial was simply integrated in the routine feedback according to the assignment mentioned above, without informing the family physicians about the trial, thus ensuring absolute blinding. Also, by contract all family physicians in the Maastricht region agreed to cooperate with our center, including with regard to research.

A blind assessment of rationality by the expert peer was not possible for practical reasons.

Rationality of requests of the tests involved in the trial was determined by relating request data to clinical data provided on every request form and comparing this information with regional guidelines or standards of the Dutch College of General Practitioners NHG for good diagnostic behavior concerning the tests involved in the trial.¹⁴

Since 1981 the Diagnostic Coordinating Center Maastricht has published guidelines on a variety of tests for primary care in its region. These guidelines are compiled by expert family physicians in collaboration with specialists, and are then authorized and distributed to all family physicians in the region.¹⁵ Such authorized guidelines were also compiled for the tests involved in the trial. These were not sent to the family physicians, but only used by the Diagnostic Coordinating Center Maastricht to determine the rationality of requests until the trial was stopped.

A request was considered to be rational when it was in accordance with guidelines or standards. If not in accordance it was considered to be non-rational, unless the request was at the insistence of the patient.

Analysis

1. Volume data

For calculating the volume of requests, data from every request for the tests involved in the trial over the period July 1988 through December 1991 were collected and put into a computer database. Data from 1 year prior to the trial (July 1988-June 1989) were used as baseline measurement and were compared with data from the first year (1990) and the second year (1991) of the trial. Data were analyzed at two different levels. First, each intervention group was compared to each control group for the two groups of tests. Second, similar comparisons

were performed for several individual tests for which a particular effect was expected before the trial. Tests that were considered to be requested unnecessarily and inappropriately for a substantial number of clearly defined indications before the trial were cervical smears and X-rays of the lumbar spine. Therefore, a decrease in the number of requests was expected especially for these tests.

The differences between the baseline measurement and the last year of the trial for each intervention group and control group were compared by means of the Mann-Whitney test.

2. Rationality data

Rationality data were only obtained from all requests of two selected months per year since assessing rationality for all requests of every month of the whole study period was too laborious. Data from March and September 1989 counted as baseline measurement. They were compared with data from March and October 1990 (first year of the experiment), April and November 1991 (second year of the experiment) and May 1992 (final measurement). Each request was compared with the guidelines and classified as rational or non-rational. In the feedback, emphasis is put on the advice to refrain from non-rational requests in future. Hence, we expected a decrease in the number of requests classified as non-rational.

Rationality data were also analyzed on test-group level and on individual test level analogous to the volume data. Before the trial, it appeared that cervical smears and X-rays of the chest and lumbar spine were frequently requested non-rationally. Therefore we expected rationality to improve for these 3 tests. In the rationality data it was possible that, due to holidays or illness for an individual family physician there were no requests in a month that was to be discussed in a feedback report. These were considered as missing data.

Data on test-group level were analyzed by means of a repeated measures analysis of variance model capable of correcting for missing data.^{16,17} For each family physician we calculated the mean number of rational and non-rational requests per test-group per monthly analysis. To improve the normality of the data, a square root transformation was used. Testing was done with the Wald test. For individual tests, where the numbers of requests were too small for this analysis, the rationality was evaluated with the Mann-Whitney test. Here the number of rational and non-rational requests was calculated per family physician per test per monthly analysis.

For all analyses, on volume as well as on rationality, a two-sided significance level of 0.05 was used.

Results

The numbers of request forms that were analyzed are depicted in table 2.

Table 2. Numbers of analyzed request forms

Quantity data: all requests per year

Quality data: all requests from two month per year

Test	quantity data			quality data		
	before	1 st year	2 nd year	before	1 st year	2 nd year
Group A:						
Electrocardiography	1764	1924	1935	305	383	321
Endoscopy	1012	1447	1449	150	167	175
Cervical smears	5444	4804	4263	897	861	719
Allergy tests	764	990	944	61	127	110
Total:	8984	9165	8591	1413	1538	1325
Group B:						
X - rays:						
. Chest	2730	2787	2526	498	549	455
. Cervical spine	560	619	567	109	108	93
. Thoracic spine	298	347	348	59	54	50
. Lumbar spine	1128	1212	1127	215	223	187
. Pelvis/hip	664	753	844	144	152	139
. Knee joint(s)	798	824	715	127	132	99
. Ankle joint(s)	340	394	354	67	80	53
. Sinuses	662	695	559	110	167	93
Ultrasound:						
. Liver/biliary tract	584	663	633	93	116	65
. Kidneys	128	215	252	67	65	49
Total:	7892	8509	7925	1489	1646	1283

For the effects of feedback on the number of requests, first the change in the mean number of tests per family physician per year ordered per group of tests is analyzed (table 3).

In the intervention group the numbers of tests from test-group A decreased by 1% after one year and by 7% after two years whereas in the control group it increased by 14% after one year and by 13% after two years. The difference between the intervention group and the control group was significant ($p=0.04$). The numbers of tests ordered from test-group B decreased ultimately in the intervention group (4%) but they also decreased in the control group. The difference between the two groups of family physicians was not significant ($p=0.1$).

Table 3. Mean numbers of requests per family physician for tests from test-group A and test-group B

	Test-group A		Test-group B	
	int.	contr.	int.	contr.
before trial	110	125	106	100
1 st year trial	109	143	113	103
2 nd year trial	102	141	101	100
p	0.04		0.1	

int. = intervention group /contr. = control group

The numbers of requests for some individual tests are elaborated in table 4. The number of requests of cervical smears in the intervention group decreased by 37% after two years. Although this number also decreased in the control group (10% after two years) the difference between the intervention group and the control group was significant ($p=0.01$). For X-rays of the lumbar spine the number of requests decreased ultimately by 6% in the intervention group while in the control group it increased by 7% ($p=0.02$).

Table 4. Mean numbers of requests per family physician for cervical smears and X-rays of the lumbar spine

	Cervical smears		Lumbar spine X-ray	
	int.	contr.	int.	contr.
before trial	62	67	14	12
1 st year trial	46	64	15	13
2 nd year trial	39	60	13	14
p	0.01		0.02	

int. = intervention group /contr. = control group

The effects of the feedback on the rationality of the requests were at first analyzed for the two separate test-groups (table 5).

Table 5. Mean number of rational and non-rational requests per family physician from test-group A and test-group B before (March and September 1989) and during the trial (March and October 1990, April and November 1991, May 1992).

Month	Rational requests				Non-rational requests			
	Test-group A		Test-group B		Test-group A		Test-group B	
	int.	contr.	int.	contr.	int.	contr.	int.	contr.
March '89	5.13	5.87	3.93	3.14	3.48	4.06	6.61	6.04
Sept. '89	5.06	5.18	2.50	3.15	3.70	3.53	5.07	6.26
March '90	4.86	6.21	5.21	3.63	4.67	4.24	5.42	6.05
Oct. '90	5.54	6.32	4.71	3.04	3.44	4.03	6.02	6.54
Apr. '91	4.55	5.47	3.40	2.69	2.18	3.29	3.71	4.81
Nov. '91	5.26	6.61	3.90	3.27	2.84	3.71	4.47	5.62
May '92	4.60	4.62	2.87	3.35	1.92	3.01	3.77	5.71
p	0.16		0.06		0.08		0.04	

int. = intervention group / contr. = control group

The trends for rational requests are relatively stable for both test-groups. With regard to the mean number of rational requests per physician the difference between the intervention group and the control group is not significant for test-group A ($p=0.16$) or test-group B ($p=0.06$). The trends for non-rational requests show a decline for test-group A, but the strongest decline is seen for test-group B. Considering the decrease in the mean number of non-rational requests per physician for test-group A the difference between the intervention group and the control group is also not significant ($p=0.08$). For tests from test-group B the decrease in the mean number of non-rational requests per physician in the intervention group is significantly greater than that in the control group ($p=0.04$).

The analyses of the effects on rationality on group level were completed by analyses of individual tests, namely for requests of cervical smears, X-rays of the chest and the lumbar spine. Table 6 shows that the number of rational requests has increased for X-rays of the lumbar spine.

The differences between the intervention group and the control group are, however, not significant ($p=0.1$). The numbers of rational requests of cervical smears and chest X-rays hardly changed in both the intervention group and the control group.

Table 6. Mean number of rational and non-rational requests per family physician before (March and September 1989 and during the trial (March and October 1990, April and November 1991, May 1992)) for selected tests discussed in the trial

CS = Cervical Smear

XCh = Chest X-ray

XLS = Lumbar spine X-ray

Month	Rational requests						Non-rational requests					
	CS		XCh		XLS		CS		XCh		XLS	
	int.	contr.	int.	contr.	int.	contr.	int.	contr.	int.	contr.	int.	contr.
March '89	4.77	4.73	2.11	1.62	0.40	0.33	2.73	2.91	2.69	2.21	1.92	1.38
Sept. '89	5.56	4.03	1.05	1.21	0.55	0.52	2.80	2.30	2.58	1.82	1.38	1.54
March '90	3.82	4.97	2.17	1.51	0.86	0.71	3.48	2.77	2.22	2.29	0.96	1.39
Oct. '90	4.24	4.80	1.62	1.22	0.96	0.48	2.14	2.64	2.29	2.88	1.44	1.22
April '91	3.40	3.46	1.65	1.78	0.77	0.42	1.39	2.25	1.64	1.28	0.91	1.65
Nov. '91	4.67	5.32	1.87	1.36	0.75	0.55	1.67	2.13	2.17	2.17	1.04	1.24
May '92	3.67	3.86	1.72	1.78	0.84	0.62	1.29	2.39	1.32	1.94	1.21	1.67
p	0.42		0.91		0.10		0.65		0.29		0.004	

int. = intervention group /contr. = control group

The number of non-rational requests of X-rays of the lumbar spine decreased in the intervention group. Compared with the control group the differences were significant ($p=0.004$).

Also the number of non-rational requests of chest X-rays decreased, but no significant differences were found between the intervention group and the control group ($p=0.29$). For cervical smears, the number of non-rational requests decreased slightly in the second year of the trial, in both the intervention group and the control group ($p=0.65$).

Discussion

In general, the feedback provided in the trial appears to have influenced both the quantity and quality of the test ordering behavior regarding tests discussed in the feedback. The results show that the number of requests of test-group A has decreased, cervical smears in particular. For test-group B the numbers of requests did not change significantly, with the exception of X-rays of the lumbar spine. Significant changes in the rationality could be found for test-group B where the number of non-rational requests decreased. On test level significant changes in the rationality were found for X-rays of the lumbar spine. However, before drawing any conclusion the following aspects need further consideration.

In our study, a routine health care activity applied for more than 4 years prior to the study was evaluated in a randomized clinical trial. This necessitated several adjustments. The most important restriction was that during the trial we had to choose tests that were to be discussed in the feedback for the first time. Ideally we would have discussed tests in the experimental feedback that were subject to the least rational requests of all. These tests had however already been discussed in the feedback prior to the experiment.

A trial concerning routine health care has also advantages. Automatically, all regional family physicians were involved in the trial. This rules out the danger of selection bias.

An important prerequisite for the success of the trial is the (single) blindness of the intervention. The addition of the experimental feedback to the routine feedback on blood tests, urine tests etcetera, made it possible to perform the trial without informed consent. During and after the trial there was no sign whatever that one or more physicians suspected the presence of a trial on the feedback provided by the Diagnostic Coordinating Center Maastricht.

The results we obtained are not uniform. While the number of tests decreased especially for tests from test-group A, rationality predominantly improved for tests from test-group B. The nature of the tests may account for these differences: Test-group B consists of radiologic and ultrasonographic tests. From separate data it appears that these are frequently requested at the insistence of patients. This makes it more difficult for a family physician to follow the guidelines in the feedback and refrain from a request. The same is true for electrocardiography, assigned to test-group A. Our data showed that requests for several tests from test-group A were already rather rational before the trial, endoscopy in particular. Therefore, the feedback could only cause a minor improvement of the rationality.

Not all tests could be discussed in the feedback with the same intensity. Also, tests are not equally sensitive to feedback. As could be expected, not all tests responded to the feedback in the same way. The strongest decrease in the number of requests was seen for cervical smears. Before the trial cervical smears were mostly requested as a screening procedure, especially for women younger than 35 years, which is, in the Netherlands, the age-group with the lowest prior probability of cervical cancer. In accordance with the standards of the Dutch College of General Practitioners, in the feedback requests for cervical smears for screening women younger than 35 years were discouraged. Indeed the reduction was particularly present in this age-group.

For the analyses of rationality, the trends in table 5 are rather freakish, due to variation from month to month. Since rationality is analyzed by requests per month, patterns may be affected by seasonal or even monthly changes in the numbers of requests. In March and November for instance, there are always many requests. These patterns are present however in the intervention group as well as the control group and therefore do not affect the analyses of differences.

The improvement of the rationality for test-group B as a whole is not reflected in every individual test analyzed. This is probably due to the fact that the number of requests for most individual tests of which the rationality could be determined was very low. The limited numbers for the analyses of the rationality are mainly caused by the fact that a two month sample in each year was taken. The determination of the rationality was too laborious to be possible for all requests.

The effects achieved in the trial are smaller than the effects which the feedback seemed to have had in the first years (1985-1988).¹² This is not surprising since in that period the feedback was not only accompanied by a reduction in the number of requests of tests discussed but also (though to a lesser extent) by a reduction of

the tests not discussed, probably due to a general learning effect. It is more difficult to achieve a reduction for tests that have already been reduced prior to the experiment.

In literature it is frequently stated that feedback is more effective when it is personal and provided by a respected colleague.^{1,2} Probably feedback has to be repeated to maintain an effect.¹⁸ Our feedback bears these features. This trial shows that the routine feedback is indeed effective in reducing the volume and improving the rationality of test requests. The extent of the effect can be assessed from the results of an observational study, performed before the trial. In that study reductions up to 95% were noticed and the feedback resulted in a substantial reduction in the total number of requests, associated with major savings in costs.¹²

It may be concluded that routine individual feedback, discussing the quantity and quality of test requests, is capable of decreasing the number of tests ordered and improving the rationality of test ordering behavior.

References

1. Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *Br Med J* 1991; 303: 398-402.
2. Buntinx F, Winkens RAG, Grol RPTM, Knottnerus JA. Influencing diagnostic and preventive performance in ambulatory care by feedback and reminders. A Review. *Fam Pract* 1993; 10: 219-28.
3. Axt-Adam P, van der Wouden JC, van der Does E. Influencing behavior of physicians ordering laboratory tests: a literature study. *Med Care* 1993; 31: 784-94.
4. Fowkes FGR, Evans KT, Hartley G, Nolan DJ, Roberts CJ, Davies ER, et al. Multicentre trial of four strategies to reduce use of a radiological test. *Lancet* 1986; i: 367-70.
5. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med* 1990; 322: 1499-504.
6. Martin AR, Wolf MA, Thibodeau LA, Dzau V, Braunwald E. A trial of two strategies to modify test-ordering behavior of medical residents. *N Engl J Med* 1980; 303: 1330-6.
7. Berwick DM, Coltin KL. Feedback reduces test use in a Health Maintenance Organization. *JAMA* 1986; 255: 1450-4.

8. Marton KI, Tul V, Sox HC. Modifying test-ordering behavior in the outpatient medical clinic. A controlled trial of two educational interventions. *Arch Intern Med* 1985; 145: 816-21.
9. Tierney WM, McDonald CJ, Martin DK, Hui SL, Rogers MP. Computerized display of past test results. Effect on outpatient testing. *Ann Intern Med* 1987; 107: 569-74.
10. Dowling PT, Alfonsi G, Brown MI, Culpepper L. An education program to reduce unnecessary laboratory tests by residents. *Acad Med* 1989; 64: 410-2.
11. Schectman JM, Elinsky EG, Pawlson LG. Effect of education and feedback on thyroid function testing strategies of primary care clinicians. *Arch Intern Med* 1991; 151: 2163-6.
12. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus JA. Effect of feedback on test ordering behaviour of general practitioners. *Br Med J* 1992; 304: 1093-6.
13. Pop P, Winkens RAG. A diagnostic centre for general practitioners: results of individual feedback on diagnostic actions. *J R Coll Gen Pract* 1989; 39: 507-8.
14. Rutten GEHM, Thomas S. NHG-Standaarden voor de huisarts. Utrecht: Bunge, 1993.
15. Beusmans GHMI, Does RJMM. Werkafspraken in het Diagnostisch Centrum Maastricht. *Huisarts Wet* 1988; 31: 154-8.
16. Dixon WJ, Brown MB, Engelman L, Jennrich RI.: BMDP Statistical software manual; to accompany the 1990 software release. Berkeley: University of California Press, 1990.
17. Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 1986; 42: 805-20.
18. Eisenberg JM. Physician utilization. The state of research about physicians' practice patterns. *Med Care* 1985; 23: 461-83.

Long-term effects of routine individual feedback on diagnostic actions

- effects over a 7-year period -

R.A.G. Winkens¹, P. Pop¹, R.P.T.M. Grol², A.M.A. Bugter-Maessen¹, A.D.M. Kester³, G.H.M.I. Beusmans⁴ and J.A. Knottnerus⁴

¹ Diagnostic Coordinating Center Maastricht

² Center for Research on Quality in Family Practice, Universities of Nijmegen and Maastricht

³ Department of Methodology and Statistics, University of Limburg, Maastricht

⁴ Department of Family Practice, University of Limburg, Maastricht

Submitted for publication

Abstract

Objective

To assess the long-term effects of routine individual feedback on test ordering behavior of family physicians.

Design

Cohort analytic study in primary care setting. Request patterns were followed for 7 years. Trends of tests discussed in feedback were compared with trends of tests not discussed. In addition, trends of tests for which an alternative was recommended were compared with trends of tests without an alternative.

Setting

A diagnostic center receiving all test requests from regional primary care physicians.

Participants

All 85 family physicians affiliated to the Diagnostic Coordinating Center Maastricht. The total patient population is now 187000.

Intervention

Routine feedback on volume and appropriateness of individually requested tests has been provided twice each year since 1985 to all family physicians.

Main Outcome Measures

Numbers of requests ordered by family physicians per year and per test or group of tests.

Results

The volume of tests decreased since 1985 and was still decreasing in 1991, after 7 years of feedback. For individual tests reductions of 32-96% were achieved, especially when an alternative was suggested. Not a single test that was discussed showed an increase. Compared to the expected trend without feedback, tests that were discussed declined by 62% after 7 years ($p < 0.001$).

Conclusions

When feedback is repeated as a routine procedure, it is capable of achieving a persistent change in test ordering behavior.

Introduction

In the last few decades, feedback has become an increasingly popular tool to change test ordering behavior. Many studies have been performed and several review articles have already summarized the findings.¹⁻⁵ These were mostly positive: feedback can influence physicians' test ordering behavior. Although it is difficult to compare the different studies, the effectiveness of feedback seems to increase when the feedback is profound, personal and provided by an expert peer.³ Also, repeating feedback is proposed.^{4,6}

In virtually all situations, feedback was provided for the purpose of a specific study, frequently in an experimental setting. A study setting does not automatically represent the usual situation. For specific studies, physicians are normally asked to participate (with the risk of selection bias) and they are mostly instructed (with the risk of a Hawthorne effect). Therefore, ideally, feedback should be a routine procedure.

An important benefit of the routine provision of feedback is the possibility of achieving long-term effects. Most studies performed to date have been of limited duration and long-term effects have seldom been found. If followed, the effect decreased soon after the feedback was stopped.⁷ So far little is known about the effects of feedback when it is given continuously or repeatedly for many years.⁶ Literature reveals only a few studies in which feedback has been used for 3½-4 years, with promising results.⁸⁻¹⁰

Since 1985, the Diagnostic Coordinating Center Maastricht has been providing feedback as a routine health care activity.¹¹ Soon after the start of the feedback, the numbers of requests decreased.⁹ Since feedback was provided every year, long-term effects could be studied. We were interested in the duration of the effect.

Not all test requests will respond to the feedback to the same extent. We expected the strongest and most persistent effects on tests that were not only advised against in the feedback but for which also an alternative procedure or test was suggested. Frequently, tests are requested at the insistence of patients. When the requested test is one that is advised against in the feedback, the physician can be in a dilemma. Also, refraining from diagnostic testing can lead to a feeling of uncertainty, and the physician may still consider (other) diagnostic testing. The availability of alternative tests can increase the willingness to follow the recommendations.

In our study, the following questions were addressed:

- What are the long-term effects of repeated routine feedback on the total number of requests and on the number of requests for individual tests? How does the trend compare with the trend without feedback?
- When does the routine feedback have its maximal effect? Does this persist or gradually diminish?
- What is the extent and persistence of the effect of feedback when an alternative is provided for tests for which requests are discouraged?

Methods

Background

The Diagnostic Coordinating Center Maastricht coordinates all diagnostic requests from 85 family physicians and serves a catchment population of, currently, 187000. Feedback on diagnostic behavior is given twice a year to all affiliated family physicians. In the feedback comments are made about inappropriate requests and recommendations are offered for rational diagnostic testing. The feedback is based on analyses of request forms filled in for every patient. As a result the feedback concerns real cases from daily practice. Requests submitted in one month are discussed with regard to quantity as well as quality of the requests. Rationality of tests ordered can be assessed because the request forms contain clinical data on the patient (history, physical signs and findings, suspected diagnosis, reason for request, etc.). In the feedback reports, patients are mentioned by name and date of birth.

Study design

We studied the long-term effects of feedback on different levels.

First, we studied the number of requests in total and for several types of tests (clinical chemistry, hematology, serology, urine and feces tests and microbiology). Request data per test, per physician per year were obtained from 1983 through 1991. Due to the volume of our data from 1983 up to the first half of 1989 only a random sample of 50% of all requests was recorded. For comparability and accuracy of trends in tables and figures, these data were doubled. From July 1989 through 1991, all requests were recorded.

To express the extent of the influence of feedback on test ordering behavior over a prolonged period, the request trend in our region was compared to that in the region of a reference laboratory. Except for the feedback, this reference laboratory was comparable to our center with regard to the degree of urbanization of the catchment area and the average family practice size. For the reference laboratory, requests from all family physicians together in their region were recorded. Data were comparable for 44 tests that could be requested in both laboratories. For anonymity reasons data from the reference laboratory are indexed to the number of patients in the Maastricht region.

Second, the maximal effect and its persistence was determined by comparing the trend of tests frequently and repeatedly discussed and advised against in the feedback with the trend of all remaining tests. In the feedback, not all requested tests are discussed with the same intensity. Several tests are discussed repeatedly over the years for every individual family physician (table 2). These tests especially reveal how long routine feedback can be effective.

Third, within this group of frequently discussed tests we evaluated the surplus value of offering an alternative test or procedure when the use of a test was advised against. For several tests, alternative tests or test procedures were given in the feedback. For the diagnostic work-up of rheumatoid arthritis, the Erythrocyte Sedimentation Rate was suggested as an alternative for Rose or Latex testing. Likewise, for renal function testing creatinine was advised instead of the usually requested combination of creatinine and urea. For thyroid function testing, Thyroid Stimulating Hormone was advised instead of T4, Free T4 and T3. We compared the trend of tests for which an alternative was recommended (Urea, Rose, Latex, T4, Free T4, T3) with the trend of tests without an alternative (Hemoglobin, Packed Cell Volume, Differential Count, Erythrocyte Sedimentation Rate, Leucocyte Count, Erythrocyte Count, Lactate Dehydrogenase, Aspartate Aminotransferase, Alanine Aminotransferase). Because of large differences in the numbers of requests per year, data from the tests with an alternative were indexed to the level of the other group in 1984.

Statistical analysis

The comparison of the Maastricht data with the reference laboratory data enabled us to quantify the long-term effects of routine feedback. For each year the request numbers per test for our center and the reference laboratory were calculated. Differences per year per test were statistically tested by means of the Mann-

Whitney test. For the evaluation of the maximal effect of routine feedback and its persistence several trends per family physician in the Maastricht region were compared and tested by means of the Wilcoxon signed ranks test. The influence of the recommendation of alternative tests was analyzed by comparing the trends for tests with or without a recommended alternative. When applicable, trends per general practitioner were taken to be the average change per year as calculated by linear regression of numbers of requests versus year.¹² In all calculations, a two-sided significance level of 0.05 was used.

Results

From 1985 through 1991 there was a mean annual decrease in the Maastricht region of 4.5%, resulting in a total decrease of 28% in 1991 (figure 1).

Apart from a temporary increase in 1988 and 1989 a decrease still occurred in 1991. Decreases can be seen especially for hematology, serology, clinical chemistry and urine/feces tests (table 1). Especially these categories contain tests that are frequently discussed in the feedback. For all these categories the fewest tests were requested in 1991, with total decrease up to 60% compared to the highest numbers of requests in 1984 or 1985.

The remaining categories (radiology, endoscopy etc.) were hardly or even never discussed before 1989, and do not show the same trends of decrease (data not presented).

The extent of the effects of our feedback was assessed by comparing our data with those of a reference laboratory. While the number of requests of discussed tests decreased almost continuously in our region the corresponding number in the reference laboratory increased (mean increase 3.2% per year). The differences between the two laboratories are statistically significant ($p < 0.001$). If the trend in the Maastricht region had been the same as in the reference region (should no feedback have been provided), the number of requests in 1991 for the frequently discussed tests would have been more than twice as high. Thus, the actual reduction for these tests is 61% after 7 years.

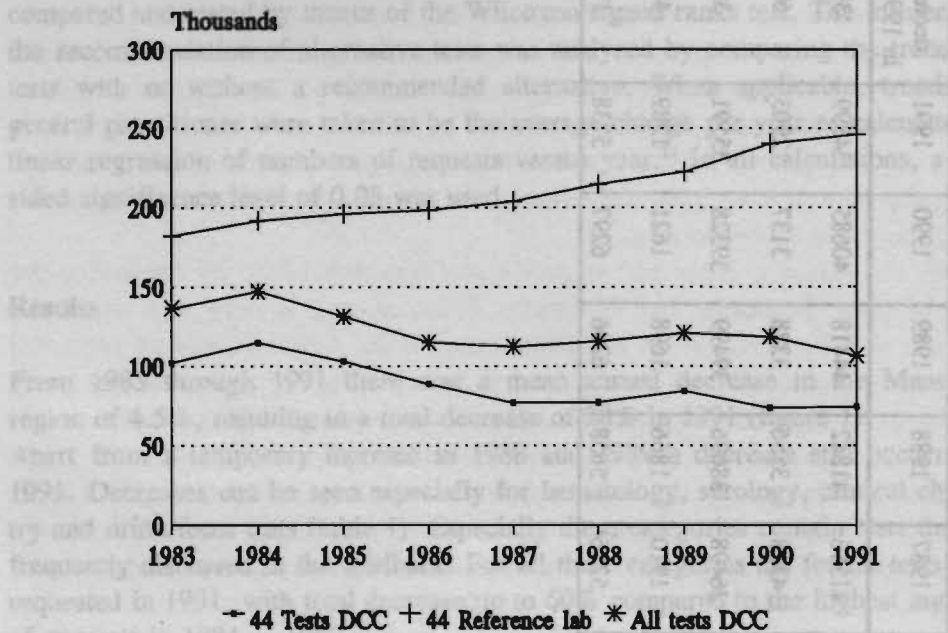
The effects of the feedback were especially seen in tests frequently discussed in the feedback in 1985 and subsequent years. Trends of these tests were compared with trends of the remaining tests (figure 2).

Table 1. Number of requests per category of tests per year; 1983 through 1991

Year Category	1983	1984	1985	1986	1987	1988	1989	1990	1991	decrease* in 1991 (%)
Hematology	53298	57344	50384	43538	41556	41842	44418	40685	34899	39
Serology	6980	7194	6262	4700	4318	3556	3328	3137	2903	60
Clinical chemistry	46456	50980	44312	37238	36498	38696	39499	39328	35591	30
Urine/Feces tests	2336	2700	2080	2068	1910	1866	1698	1621	1389	48
Bacteriology/Virology	6676	6880	4622	4916	5216	5658	6296	6297	5738	17

* Decrease in 1991 compared to the highest level in 1984

Figure 1. Trend of the total number of requests per year, including the trend for 44 tests comparable with a reference laboratory; 1983 through 1991



While the frequently discussed tests decreased almost continuously, with a total decrease of 51%, the remaining tests increased after a short period of decrease. Ultimately the trend of these infrequently discussed tests remained relatively stationary. Compared to the trend expected should no feedback have been provided, the actual reduction for frequently discussed tests was 62% after 7 years. The differences between trends are statistically significant ($p<0.001$).

Several tests are repeatedly designated in the feedback as inappropriate for specific indications for which they are often requested. For these tests table 2 shows that even after giving feedback for 7 years, the effects of the feedback on the numbers of requests persist.

Total decrease ranges from 32% up to 96%. Typically, none of the tests presented in table 2 show a substantial increase after the first effects of the feedback. Several tests, in particular all hematological tests, Alanine Aminotransferase and Aspartate Aminotransferase still decreased in 1991. Before we started evaluating the feedback our hypothesis was that family physicians would react to the feedback on a specific test more if an alternative was provided.

Table 2. Numbers of requests for several individual tests per year from 1983 through 1991

	1983	1984	1985	1986	1987	1988	1989	1990	1991	decrease* in 1991 (%)
Hemoglobin	10828	11776	10392	9244	8422	7832	8073	7811	7050	40
PCV	3354	3648	3756	3430	2946	2996	3319	2943	2469	34
ESR	12320	13300	11974	10220	10102	10052	10396	9572	8299	38
Diff. Count	9966	10834	8834	6796	5962	6140	6488	5558	4161	62
Leucocytes	10562	11196	9478	7572	6984	7282	7664	6778	5429	52
Erythrocytes	4622	4864	4316	3944	3450	3002	3216	2857	2283	53
Urea	2482	2770	1594	634	170	106	70	86	109	96
Iron/Transf.	454	410	588	356	226	228	243	260	242	59
ASAT	2074	2048	1682	1228	1236	1494	1376	1510	1216	41
ALAT	4860	5160	4152	3268	3226	3144	3051	2875	2610	49
LDH	874	852	670	420	468	406	439	481	480	44
Rose	1190	1184	972	508	440	246	175	150	148	88
Latex	1158	1130	926	484	400	242	163	138	137	88
Free T4	70	234	840	936	704	488	439	540	515	39
T4	862	856	346	32	8	14	16	49	50	94
T3	148	90	32	6	4	6	5	18	14	84
TSH*	10	4	42	38	518	1008	1486	1650	1934	**
Creatinine	4604	5166	4678	4380	4216	4504	4552	4451	3844	

Bold numbers: the lowest number of requests

* decrease in 1991 compared to highest level in 1984 or 1985

** increase in accordance with recommendations in feedback

Figure 2. Trends of the numbers of requests; frequently discussed tests versus the rest; 1983 through 1991

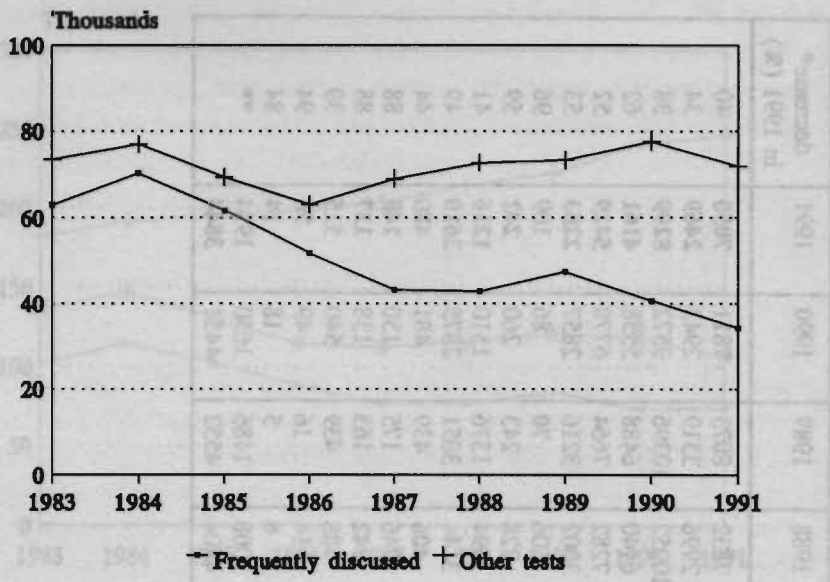
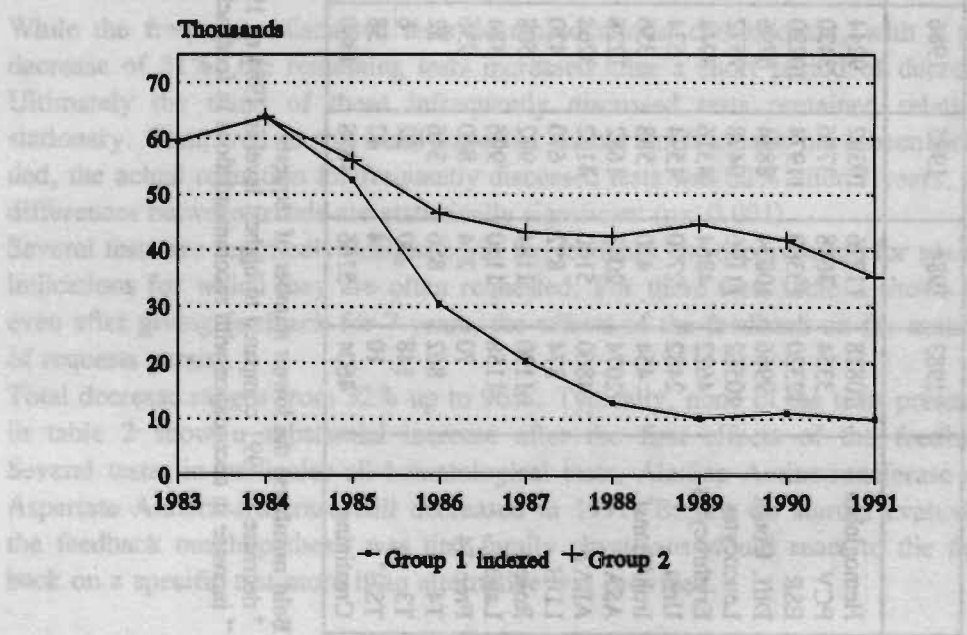


Figure 3. Trends of the numbers of requests per year for tests with (group 1) or without (group 2) an alternative recommended in the feedback



For these tests the numbers of requests have decreased almost to zero. As an alternative for Free T4, T4 and T3, Thyroid Stimulating Hormone was advised. Correspondingly, the number of Thyroid Stimulating Hormone requests increased rapidly. The combination urea and creatinine was "replaced" by requests of creatinine alone. For rheumatoid arthritis, a shift occurred from Rose and Latex to the Erythrocyte Sedimentation Rate. The trend for tests with a recommended alternative showed a considerably greater decrease than the trend for tests without an alternative (figure 3) ($p=0.002$).

Discussion:

The results show an overall reduction of the number of requests that persisted in the period 1985-1991. For specific tests ultimately reductions of 32-96% were achieved, especially for tests for which an alternative was recommended. In no situation did a relapse occur. For a correct interpretation of our findings the following aspects must be taken into consideration.

We obtained results from an observational study. An experiment was not possible. We started providing feedback in 1985 at the request of the family physicians. All of them received the feedback and nowadays the feedback is an accepted and appreciated routine activity. Simply stopping the feedback for a group of (randomly assigned) family physicians for the purpose of a study would have been unwise and unnecessary since it has already been demonstrated that an effect of feedback is not continued once feedback is stopped.⁷

We now have evidence that feedback is capable of achieving effects lasting at least 7 years. It is even more typical that the maximal effect does not yet seem to have been reached in the final year of the study period. Apparently, the strongest effects are not achieved when feedback is provided for a short period of time but when it is repeated. This stresses the necessity of the continuation of feedback as a routine procedure.

In virtually all trends that can be seen in the figures presented in this paper, there is a small increase in the number of requests in 1988 and 1989, suggesting that a maximal effect had been achieved. This proved to be incorrect. In 1990 and 1991 again a decrease was seen. Possibly the renewed decrease thereafter was caused by a more decision analytical approach in the feedback with more emphasis on the diagnostic consequences of tests. The decrease in the total number of requests in 1991 can be explained partially by an experiment held from fall 1989 through spring 1992. During this experiment feedback was given on several tests not

discussed before, among which electrocardiography, cervical smears and a diversity of radiological and ultrasonographical tests. The strongest effects of this experiment were seen in 1991.

While the numbers of requests nationwide increased every year, in the Maastricht region the overall numbers decreased since 1985. Apparently not only the tests discussed are influenced by the feedback but also those tests that are not discussed. The reason for this might be a general learning effect arising from the feedback. This general learning effect, however, seems to disappear again in 1987 and the ensuing years.

Although the request patterns are themselves evidence for a long-term effect of feedback, we tried to seek statistical confirmation. The differences we found must, however, be considered with caution, since they are reduced by the fact that the "not discussed" tests are influenced by the feedback as well due to a general learning effect. The differences with a situation in which no feedback is given, as in the reference laboratory, are presumably more clear-cut.

The numbers of requests of Free T4 do not reveal the actual trend. Of all Thyroid Stimulating Hormone requests, 23% have abnormal test results and lead to determination of Free T4 as well, regardless the fact whether it is requested by the physician or not. Especially in the recent years, this has caused the smaller decrease of the Free T4 requests. In 1990 and 1991 less than 15% of all Free T4 requests were actively requested by family physicians.

Obviously the decrease in the number of requests has reduced the costs of diagnostic testing. Based on the national trend of up to 8% increase in total costs per year, the expenditure for diagnostic tests would have been considerably higher than it is now. On the basis of nationally fixed charges per test, instead of an expected sum of 2 million US dollars in 1991 if no feedback had been given, less than 1 million US dollars was spent on diagnostic testing by family physicians. Giving feedback costs at most 85000 US dollars each year. After deducting these costs from the savings in the period 1985-1991 in our region a net sum of 3.6 million US dollars was saved. This contrasts with earlier findings of Schroeder et al, who concluded that the savings in costs of requests are outweighed by the expenditure for the feedback itself.¹³ The financial benefit of our feedback appeared to be considerable and increased as the feedback was continued. Besides the quality improvement achieved, this signals another important reason for continuing the feedback. Once feedback is started, the savings increase with every year. In other words, the longer you continue, the higher the return. Based on the findings presented we conclude that routine individual feedback is

capable of reducing the number of requests persistently. As long as the feedback is continued, the effects remain clearly visible. The strongest decrease is seen for tests for which an alternative is provided.

Acknowledgement

This study was supported by a grant from the Dutch Ministry of Welfare, Health and Cultural Affairs, study number 88-28.

References

1. Fraser CG, Woodford FP. Strategies to modify the test-requesting patterns of clinicians. *Ann Clin Biochem* 1987; 24: 223-31.
2. Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *Br Med J* 1991; 303: 398-402.
3. Horder J, Bosanquet N, Stocking B. Ways of influencing the behaviour of general practitioners. *J R Coll Gen Pract* 1986; 36: 517-21.
4. Buntinx F, Winkens RAG, Grol RPTM, Knottnerus JA. Influencing diagnostic and preventive performance in ambulatory care by feedback and reminders. A review. *Fam Pract* 1993; 10: 219-28.
5. Grossman RM. A review of physician cost-containment strategies for laboratory testing. *Med Care* 1983; 21: 783-801.
6. Eisenberg JM. Physician utilization. The state of research about physicians' practice patterns. *Med Care* 1985; 23: 461-83.
7. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med* 1990; 322: 1499-504.
8. Lyle CB, Bianchi RF, Harris JH, Wood ZL. Teaching cost containment to house officers at Charlotte Memorial Hospital. *J Med Educ* 1979; 54: 856-62.
9. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus JA. Effect of feedback on test ordering behaviour of general practitioners. *Br Med J* 1992; 304: 1093-6.
10. Winickoff RN, Coltin KL, Morgan MM, Buxbaum RC, Barnett GO. Improving physician performance through peer comparison feedback. *Med Care* 1984; 22: 527-34.
11. Pop P, Winkens RAG. A diagnostic centre for general practitioners: results of individual feedback on diagnostic actions. *J R Coll Gen Pract* 1989; 39: 507-8.

12. Altman DG. Practical statistics for medical research. 1st ed. London: Chapman & Hill, 1991.
13. Schroeder SA, Myers LP, McPhee SJ, Showstack JA, Simborg DW, Chapman SA, et al. The failure of physician education as a cost containment strategy. JAMA 1984; 252: 225-30.

General practitioners and diagnostic testing. Why are there differences in test ordering patterns and in the change of ordering by feedback?

In the study presented here, we tried to find factors that explained differences in test ordering between the two groups as a result of the feedback and thereby general practitioners' decision not to order a second laboratory test. In the study presented here, we tried to find factors that explained differences in test ordering between the two groups as a result of the feedback and thereby general practitioners' decision not to order a second laboratory test. In the study presented here, we tried to find factors that explained differences in test ordering between the two groups as a result of the feedback and thereby general practitioners' decision not to order a second laboratory test.

Summary

The test ordering behaviour of 85 general practitioners was changed by means of routine individual feedback. A questionnaire and interview were used to study which physician and/or practice characteristics could explain differences in test ordering behaviour and the extent to which physicians tend to change their behaviour according to the guidelines provided in the feedback.

Data from the questionnaire and interview were related to data on test requests as obtained from the Diagnostic Coordinating Centre Maastricht.

With the exception of the number of hours per week general practitioners spent in practice ($p=0.01$) no factors were found that influenced test use per general practitioner.

General practitioners with more experience ($p<0.001$) or with a shorter duration of consultations ($p=0.03$) changed more in accordance with the feedback.

Despite the findings of this study it is still not known sufficiently what factors explain why doctors do, or do not, change their behaviour according to recommendations in feedback.

Introduction

There are great differences between general practitioners with regard to the use of diagnostic tests.^{1,2} In literature, a variety of factors is held responsible for these differences between doctors. Among these factors, the attitude to risk-taking by the general practitioner, the type of practice, the amount of experience and the age of the general practitioner are most often mentioned.³⁻¹¹ Not all studies, however, arrive at a uniform opinion. Different study designs and study settings may partially explain these inconsistencies. Possibly, there are also international or even interregional differences in factors of influence.¹²

The numbers of requests are increasing every year. Many of these requests are considered to be unnecessary. Several methods have been applied to reduce the number of unnecessary requests, among which feedback.¹³ In many studies the effectiveness of feedback is studied; although effects are not guaranteed, feedback can be an effective way to change test ordering behaviour.¹⁴⁻¹⁸ Just like test ordering behaviour itself, the degree of change resulting from feedback varies among general practitioners. Possibly, factors that explain differences in test use among general practitioners can also explain differences in the response to interventions such as feedback. By response we mean the extent to which general practitioners change their test ordering behaviour and bring it in accordance with the recommendations in the feedback.

In literature no satisfactory information was found that explained why one general practitioner changes his behaviour as a result of the feedback and another general practitioner does not. It was not sufficiently known whether the factors that are mentioned in literature as explaining differences in test use could also explain differences in the response to the feedback.

In the study presented here we tried to find factors that explained differences in the response to feedback on test requests. Also, by means of the same data, we tried to find factors that explained differences in test use by general practitioners.

Our aim was to find factors that sufficiently explained why general practitioners respond or, even more important, do not respond to the feedback. Then, those general practitioners could be identified who need feedback the most or to whom it should be repeated. Also it would then be possible to put more emphasis on subgroups of general practitioners who respond poorly to the feedback. Perhaps then for these subgroups the response to the feedback could be improved in the future.

In the study the following questions were addressed:

1. What factors explain variations among general practitioners with regard to test use?
2. What factors explain why a general practitioner does or does not change test ordering behaviour in accordance with the recommendations in the feedback?
3. Can general practitioners reliably predict whether their test ordering behaviour has changed due to feedback?

Methods

Background

The Diagnostic Coordinating Centre Maastricht coordinates (all) diagnostic test requests from the 85 general practitioners in the catchment area, serving a population of now 187000 inhabitants. Since 1985, individual feedback has been provided twice a year to every individual general practitioner who regularly requests tests at our centre. The feedback tries to improve the rationality of test requests and to reduce the number of unnecessary requests.¹⁹

The start of the feedback was followed by a reduction in the volume of tests. Especially for tests that were frequently designated as inappropriate for the indication for which they were requested the numbers decreased dramatically, with decreases up to 95% after two years.²⁰ Not all general practitioners, however, responded to the feedback to the same extent and differences between general practitioners are large.

The study design

In our study, request data were related to data concerning practitioner and practice characteristics, as well as to the opinion about our centre and the feedback that had been provided in the past years.

On average, 85 general practitioners work in practice and order tests at our centre. General practitioners were included if they were working in practice at least one year before the study.

Data collection

Data on practitioner and practice characteristics were obtained by means of structured interviews and a mailed questionnaire among all included general practitioners. The interview was the most reliable way to obtain information; also it minimized the non-response. The interviews were taken by a trained, independent interviewer. To obtain reliable answers, delicate questions, such as on practice size, were asked in a mailed questionnaire. In a pilot three general practitioners who for several reasons (such as retirement) could not be enrolled in the study, were interviewed.

In the interview and the questionnaire the following aspects were considered:

- Practitioner characteristics: education (vocational training yes/no), amount of experience (years in practice), mean duration of consultations, conception of one's task (a list of 21 items), attitude towards and use of guidelines or standards of care (open questions), attitude to risk-taking (a list of 5 items), attitude towards the use of diagnostic tests (a list of 5 items), participation in continuing medical education (CME) (a list of 6 items, including the perception of the influence of CME), keeping up with literature (especially on diagnostic tests in Dutch journals) (2 items, including the perception of its influence), indulging in medical hobbies (specific fields of medical interest)(open questions).

The attitude to risk-taking and the conception of the doctors' task were assessed by means of validated standard lists of questions from other studies.²¹

- Practice characteristics: type of practice (solo, group or health centre), practice size (number of patients), manpower in practice (including fulltime equivalents), participation in scientific studies in which tests need(ed) to be requested (open questions), diagnostic facilities in the own practice (a list of 24 possibilities).
- The Diagnostic Coordinating Centre Maastricht: satisfaction with our centre (a list of 6 items) and its feedback (4 lists with a total of 23 items).
- Estimation of the perception of the practitioner's own changes in test ordering behaviour due to the feedback (one multiple choice question).

In total the list of questions for the interview contained 29 questions of which 15 questions consisted of up to 23 items. The mailed questionnaire contained 9 short questions.

Request data were obtained from the Maastricht diagnostic centre. Data on the total number of tests per general practitioner in the period July 1990 until June 1991 (one year prior to the interviews) were used as a measure for test ordering

behaviour. For the determination of the response to the feedback similar request data were used. The response was expressed as the change in terms of percentage in the number of requests for 11 selected tests. This change was determined by comparing the number of requests before a general practitioner received his first feedback report with the number of requests prior to the study (in 1990). The difference was expressed as a percentage of the corresponding number of requests in the year before the first feedback report.

Probably, the response to the feedback depends primarily on test ordering behaviour itself. Those who request few tests may need smaller changes as a result of feedback. This factor might explain most of the differences found in the analyses hereafter. Therefore, in the analyses concerning the response to feedback we corrected for differences in test ordering behaviour by determining the change in the numbers of requests as a percentage of the numbers of requests prior to the first feedback. The 11 selected tests were frequently and repeatedly discussed in the feedback for all general practitioners since the start of the feedback in 1985. In the feedback it was recommended to refrain from the use of these tests or to request them only for specific occasions. A change in the request patterns was especially seen for these 11 tests.²⁰

The analysis

Per general practitioner data from the interview and the questionnaire were related to data on the test ordering behaviour and the response to the feedback. Data on the numbers of requests had a non-normal and skewed distribution. This was remedied using a log-transformation. The same log-transformation was applied on the data on the response to the feedback too. Using multiple linear regression, variables from the interview or questionnaire were selected that had a significant contribution to the explanation of the differences in test ordering behaviour and the response to the feedback.

Results

Of all general practitioners in our region, 6 were excluded since they had been working less than 1 year in our region or had retired from practice recently. One general practitioner was excluded since he was involved in the development of the questionnaire. As a result, 78 general practitioners were asked to participate.

Three of these 78 general practitioners refused, resulting in a non-response of 4%. The test ordering behaviour could be determined for all 75 responders to the interview and the mailed questionnaire. The response to the feedback, however, could only be determined for 67 general practitioners since for 8 of the 75 participants only insufficient data could be obtained on the numbers of requests before they received feedback for the first time.

Obviously, the participants were not a random selection from the total population of general practitioners. With regard to several characteristics the participants were compared with the average Dutch general practitioner. Table 1 shows that the general practitioners in the Maastricht region do not differ substantially from the whole population of Dutch general practitioners.

Table 1. Characteristics of the participating general practitioners compared to all Dutch general practitioners

Variable	Maastricht region	Nationwide*
N	75	6595
mean age	?	44**
mean years in practice	15	?
gender		
* male (%)	89	85
* female (%)	11	15
practice setting		
* solo (%)	51	52
* group practice (%)	40	39
* health centre (%)	9	9
urbanization		
* city	68	25
* others	32	75
mean number of patients	2250*	2310

* Source: Data from general practitioners registrations, NIVEL 1993, Utrecht

** Estimation

All responders completed the questionnaire and participated in the interview. Most questions were answered completely, with the exception of those regarding practice size and the open questions on specific fields of medical interest or hobbies, continuing medical education and guidelines that are actively implemented in the practitioner's own practice.

For several questions the answers were identical in more than 80% of cases and therefore could not sufficiently discriminate subgroups of general practitioners.

This concerned questions regarding the opinion about our centre and the feedback that had been provided in the past. Almost all responders indicated to be satisfied about the diagnostic centre and its feedback. The results of the analyses are summarized in the undermentioned tables 2, 3 and 4.

Table 2. Results from multiple linear regression for practitioner characteristics

For non-significant variables the p-value for the partial relation is given, corrected for the significant variables

Dependent variables:

1. number of requests per practitioner
2. change in the number of requests for 11 frequently discussed tests

Variable	Test ordering behaviour		Response to feedback	
	partial correlation	p	partial correlation	p
vocational training	0.10	0.45	0.04	0.86
experience	0.12	0.40	0.47	0.001
duration of consultations	0.08	0.57	0.29	0.03
conception of task	0.10	0.46	0.09	0.62
use of guidelines	0.19	0.42	0.29	0.18
attitude to risk-taking	0.03	0.81	0.15	0.31
attitude to test use	0.09	0.57	0.12	0.35
participation in CME	0.22	0.47	0.31	0.16
reading literature	0.13	0.64	0.20	0.35
specific medical interest	0.11	0.59	0.07	0.79

Table 3. Results from multiple linear regression for practice characteristics

For non-significant variables the p-value for the partial relation is given, corrected for the significant variables

Dependent variables:

1. number of requests per practitioner
2. change in the number of requests for 11 frequently discussed tests

Variable	Test ordering behaviour		Response to feedback	
	partial correlation	p	partial correlation	p
practice type	0.09	0.46	0.23	0.06
manpower	0.32	0.01	0.06	0.73
participation in studies	0.25	0.07	0.14	0.37
use of own test facilities	0.01	0.98	0.01	0.96

Table 4. Results from multiple linear regression for the view at the DCC and its feedback including the perception of changes
For non-significant variables the p-value for the partial relation is given, corrected for the significant variables

Dependent variables:

1. number of requests per practitioner
2. change in the number of requests for 11 frequently discussed tests

Variable	Test ordering behaviour		Response to feedback	
	partial correlation	p	partial correlation	p
satisfaction with DCC	0.16	0.17	0.11	0.35
satisfaction with feedback	0.18	0.15	0.17	0.21
use of feedback	0.11	0.37	0.18	0.19
changes through feedback	0.28	0.02	0.13	0.29

Test ordering behaviour was related in the first place to practice size. Sizes ranged from 1800 to 6900 (the latter being the total of group practices) and were not supplied in 9 cases. Furthermore the type of practice could be important, since this is related to the effective number of patients for a particular general practitioner. For solo practices, there was an interaction with the type of practice, resulting in an increase of log (requests) of 0.647 per 1000 patients.

For other practice types (group practices and health centres), the increase was not significantly different from zero. Moreover, a significant positive relation with the hours per week in practice (manpower) was found ($p=0.01$). Other variables had no significant relation to the number of requests. For example, we found no significant relation with risk-taking nor conception of task ($p=0.81$ and $p=0.46$ respectively). In the analyses there were two significant outliers (tested by Bonferroni-corrected outlier test).²²

As expected, the response to feedback was closely related to test ordering behaviour itself ($p<0.0001$). For the response to feedback, a relation with practice size and type of practice was not found, but relations with the mean duration of the consultations and the number of years of experience were significant. Experience and mean duration of consultations correlate very well with each other ($p<0.001$). When a doctor is more experienced, consultations take less time. In the analysis of the two factors, they are used as mutual co-variables. There was a strong relation of response (log) to the years of experience ($p<0.001$). Quantitatively the response quotient decreases about 5% for each year of experience. Also, although

somewhat less strong, a relation was found for the mean duration of the consultations ($p=0.03$). General practitioners who work faster apparently respond better to the feedback. Other variables such as attitude to risk-taking and conception of task did not show any significant relation with a change in test ordering behaviour as a response to the feedback ($p=0.35$ and $p=0.62$ respectively). Likewise, we found no evidence for a co-effect from literature or continuing medical education. We were especially interested in the perception of the general practitioners with regard to their own change in test ordering behaviour due to the feedback. Does it have a predictive value for an actual change? This perception was not significantly related to an actual change ($p=0.29$). Typically, however, the perception was significantly related to the test ordering behaviour itself ($p=0.02$).

Discussion

The results show that, while the non-response was low and by far the majority of the responders were satisfied with our centre and the feedback, only few characteristics can explain differences in test ordering behaviour or the response to feedback. Before drawing any conclusion, the following aspects need further consideration.

The most reliable answers to our questions in the interview and the questionnaire would be obtained if they were anonymous. This, however, was not possible since we had to couple the answers to the request data from each individual general practitioner.

The response to the study was good. Possibly, this is associated with the generally positive opinion about the centre and its feedback. This satisfaction had already been expressed at other occasions prior to our study. The 3 non-responders all refused because of their dissatisfaction with our centre, the feedback in particular. Although the high satisfaction in itself is a pleasant finding, the uniformity makes it impossible to study factors related to satisfaction.

An important practice characteristic was practice size. Unfortunately, the answers to the corresponding questions are not totally reliable. Practice size is a touchy subject and data are at times provided reluctantly. Apart from the great number of questionnaires in which practice size was not given (9 cases, 12%), a few general practitioners afterwards admitted that they filled in an incorrect number of patients. Consequently, our data on practice size were possibly unreliable.

Therefore, the outcome of the analyses on practice size must be considered with caution.

As in other studies we found only a few characteristics that are related to differences in test ordering behaviour. In contrast with findings of others we found no influence of factors like diagnostic facilities in practice, attitude to risk-taking and type of practice. This is particularly surprising concerning the attitude to risk-taking since in our study the same questions were used as in the studies published earlier.^{6,21} Concerning the other aspects (diagnostic facilities in practice and type of practice) differences in the study findings may be explained by differences in the questionnaires, study design or target group.

It appears from our findings that the more experienced a general practitioner gets, the more he is willing to act according to the recommendations in the feedback. This is somewhat contradictory to the viewpoint of Metcalfe who stated that especially younger doctors have a positive attitude towards feedback.²³ It is generally accepted that feedback (or any other intervention) can only be effective when it is accepted by the target group. Consequently, one might expect especially younger general practitioners to respond to the feedback. Our study shows that the opposite is true. Based on this finding it may be wise in future to put more emphasis on feedback to younger, less experienced general practitioners.

The shorter the duration of a consultation, the better the response to the feedback. Apparently general practitioners tend to follow the recommendations better under a higher work load. This finding is surprising. It would seem more obvious for general practitioners to revert to the same old habits when there is a lack of time.

We found no factors that explained both variations in test use as well as in the response to the feedback. Apparently factors that influence test use do not predict whether, and to what extent, a general practitioner will respond to feedback. Test ordering behaviour itself correlated well with the response to the feedback. Possibly the general practitioners who requested many tests subsequently received more extensive feedback reports and are therefore possibly influenced stronger.

The effects of feedback depend on the personality of the reviewer and on the degree of acceptance of the feedback by the target group.¹⁷ Unfortunately this assumption could not be evaluated in our study since we did not change the reviewer or the feedback procedure nor did we compare these with reviewers and feedback in other regions.

This study has traced only few characteristics that explain differences in test ordering behaviour among general practitioners or differences in the extent in which they respond to feedback. Furthermore we conclude that there are inconsistencies with the results of other studies. Our most conspicuous finding was that more experienced general practitioners respond best to the feedback.

Test ordering behaviour is influenced by a variety of factors, but for the time being it is still not known sufficiently what factors are responsible to what extent. Further research therefore remains necessary. When explanatory factors, especially concerning the response to interventions like feedback, are known better it might be possible to optimize the methods that try to change test ordering behaviour.

References

1. Schroeder SA, Schlifftman A, Piemme TE. Variation among physicians in use of laboratory tests: relation to quality of care. *Med Care* 1974; 12: 709-13.
2. Forbes GB. The use of hospital laboratory facilities by general practitioners. *J R Coll Gen Pract* 1966; 11: 143-9.
3. Epstein AM, Begg CB, McNeil BJ. The effects of group size on test ordering for hypertensive patients. *N Engl J Med* 1983; 309: 464-8.
4. Epstein AM, Begg CB, McNeil BJ. The effects of physicians' training and personality on test ordering for ambulatory patients. *Am J Publ Health* 1984; 74: 1271-3.
5. Epstein AM, McNeil BJ. Physician characteristics and organizational factors influencing use of ambulatory tests. *Med Decis Making* 1985; 5: 401-15.
6. Zaat JOM, van Eijk JThM. General practitioners' uncertainty, risk preference, and use of laboratory tests. *Med Care* 1992; 30: 846-54.
7. Hlatky MA, Lee KL, Botvinick EH, Brundage BH. Diagnostic test use in different practice settings. *Arch Intern Med* 1983; 143: 1886-9.
8. Hartley RM, Charlton JR, Harris CM, Jarman B. Patterns of physicians' use of medical resources in ambulatory settings. *Am J Public Health* 1987; 77: 565-7.
9. Davis PB, Lay Yee R. Patterns of care and professional decision making in a New Zealand general practice sample. *N Z Med J* 1990; 103: 309-12.
10. Nightingale SD. Risk preference and laboratory use. *Med Decis Making* 1987; 7: 168-73.
11. Eisenberg JM, Nicklin D. Use of diagnostic services by physicians in community practice. *Med Care* 1981; 19: 297-309.

12. Hartley RM, Epstein AM, Harris CM, McNeil BJ. Differences in ambulatory test ordering in England and America. Role of doctors' beliefs and attitudes. *Am J Med* 1987; 82: 513-7.
13. Eisenberg JM. Physician utilization. The state of research about physicians' practice patterns. *Med Care* 1985; 23: 461-83.
14. Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice. A review. *Br Med J* 1991; 303: 398-402.
15. Buntinx F, Winkens RAG, Grol RPTM, Knottnerus JA. Influencing diagnostic and preventive performance in ambulatory care by feedback and reminders. A review. *Fam Pract* 1993; 10: 219-28.
16. Axt-Adam P, van der Wouden JC, van der Does E. Influencing behavior of physicians ordering laboratory tests: a literature study. *Med Care* 1993; 31: 784-94.
17. Horder J, Bosanquet N, Stocking B. Ways of influencing the behaviour of general practitioners. *J R Gen Pract* 1986; 36: 517-21.
18. Fraser CG, Woodford FP. Strategies to modify the test-requesting patterns of clinicians. *Ann Clin Biochem* 1987; 24: 223-31.
19. Pop P, Winkens RAG. A diagnostic centre for general practitioners: results of individual feedback on diagnostic actions. *J R Coll Gen Pract* 1989; 39: 507-8.
20. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus JA. Effect of feedback on test ordering behaviour of general practitioners. *Br Med J* 1992; 304: 1093-6.
21. Grol R, Whitfield M, Maeseneer J de, Mokkink H. Attitudes to risk taking in medical decision making among British, Dutch and Belgian general practitioners. *Br J Gen Pract* 1990; 40: 134-6.
22. Weisberg S. *Applied linear regression*. 2nd ed. New York: John Wiley & Sons. 1985.
23. Metcalfe DHH. Medical audit. *Br Med J* 1974; iii: 327.

Does a reduction in test use lead to more referrals?

R.A.G. Winkens¹, R.P.T.M. Grol², G.H.M.I. Beusmans⁴, A.D.M. Kester³, J.A. Knottnerus⁴ and P. Pop¹

¹ Diagnostic Coordinating Centre Maastricht

² Centre for Research on Quality in Family Practice, Universities of Nijmegen and Maastricht

³ Department of Methodology and Statistics, University of Limburg, Maastricht

⁴ Department of Family Practice, University of Limburg, Maastricht

Abstract

Objective

To test the hypothesis that the referrals of general practitioners would increase when their use of diagnostic tests is reduced through feedback.

Design

Comparison of changes in request and referral trends in a region where feedback on test ordering behaviour was provided.

Setting

A diagnostic centre in Maastricht giving routine feedback and a regional health insurance service.

Subjects

A selection of 64 general practitioners in the Maastricht region.

Main outcome measures

Numbers and trends of tests and first referrals by general practitioners.

Results

The reduction of requests was not accompanied by a higher referral rate, not even for specialisms related to tests discussed in feedback. Good responders to feedback even had decreased referral rates in contrast to increased rates for poor responders ($p=0.01$).

Conclusions

Reducing the volume of tests through feedback does not lead to more referrals. Together with lower test use, fewer referrals were seen.

Introduction

Expenditures for health care are increasing. Measures to achieve a more efficient use of medical care are needed. With this end in view, methods have been developed to reduce the use of diagnostic or therapeutic facilities. One of the ways to change diagnostic test ordering behaviour is through feedback.¹⁻⁴ It can reduce the number of requests leading to considerable savings in costs, although this is not a matter of course.^{5,6}

One important factor seems not to be discussed regarding savings in costs through feedback, namely the fact that refraining from one action may lead to another action as an "escape" or substitute. In primary health care such an escape is possible by referral. This can be explained as follows: fewer requests may lead to less certainty in the diagnostic work-up, at least from the applicant's viewpoint.⁷ When diagnostic testing is discouraged but yet more certainty is desired, alternatives might be considered. Also, seen from the patient's perspective, when a demand for testing is not granted, a demand for referral may occur sooner or later. Finally, the increased consumption of medical care may lead to a higher pressure of work, and thus to higher request and referral rates.⁸

What is known of the relation between test ordering behaviour and referral behaviour? Literature shows that general practitioners who request many diagnostic tests also refer patients more easily.⁸ Regarding potential explanatory factors there is no clear evidence. Concerning practice, doctor and patient characteristics there is only limited and contradictory information.⁸⁻¹¹ Hartley, for example, concluded that older doctors tend to request fewer tests but no significant correlation with referral rates was found.⁸ Wilkin found that younger doctors had fewer referrals.⁹ Mookink found that doctors with an "integrated" practice style have a patient- and goal-oriented approach and are reserved in the use of tests and referrals.¹¹

What happens when test ordering behaviour is changed? An adverse effect on referral behaviour of interventions like feedback discussing the appropriateness of diagnostic tests has not yet been studied. If fewer requests led to more referrals, feedback would possibly do more harm than good. Although such an adverse effect was theoretically possible, we did not expect it to occur. Moreover, recent research has shown that, even after direct intervention, referral behaviour itself is difficult to change.¹²

Since 1985, the Diagnostic Coordinating Centre Maastricht has been giving feedback to every general practitioner in its area. Overviews showed a reduction in the volume of requested tests. Tests frequently designated as inappropriate showed reductions up to 95%.⁵

We studied the occurrence of adverse effects on referral rates of our feedback by analysing referral patterns in our region over a period in which feedback has changed test ordering behaviour considerably. The following questions were addressed:

- What is the relation between test ordering behaviour and referral behaviour?
- Is a reduction in the volume of test requests through feedback accompanied by a higher referral rate?

Methods

The Diagnostic Coordinating Centre Maastricht serves 85 general practitioners in a catchment area of 187000. Feedback on test ordering behaviour is provided twice each year. The feedback basically aims at improving the rationality of test ordering behaviour. In the feedback comments are made about inappropriate requests and recommendations are offered with regard to rational diagnosis. The feedback is based on request forms filled in by the general practitioner in the course of one month.¹³

Data collection

Request data were available from the Diagnostic Coordinating Centre Maastricht. The number of requests per year per general practitioner per test was calculated from 1986-1991. Data on referrals per general practitioner per specialism per patient (anonymously) were obtained from the regional health insurance service VGZ. In the Netherlands the first referral is mostly at the initiative of the general practitioner, or sometimes requested by the patient. Repeated referrals are mostly not initiated by the general practitioner. Therefore, for the comparison of referral data with request data only first referrals were used. Data on first referrals were available over the period 1986-1992.

The analysis

In a retrospective analysis request data and referral data from the Maastricht region from 1986-1991 were studied. For every year per individual general practitioner the number of referrals was related to the number of requests, in order to determine the correlation between these two factors. The possibility that a lower test use through feedback affected referral behaviour was studied on three different levels.

First, for every general practitioner we studied the correlation between the request trend and the referral trend over the period 1986-1991.

Second, referrals were analyzed and compared for individual specialisms. We expected changes in referral patterns (if present) due to the feedback particularly for those closely related to the clinical problems for which diagnostic tests are frequently requested. These specialisms were internal medicine and, to a lesser extent, orthopaedics. Referral trends for these specialisms were compared to those of two specialisms not closely related to tests discussed in the feedback, namely surgery and gynaecology. Trends of first referrals from 1986-1991 were compared for these four specialisms.

Finally, referral trends were compared for two groups of general practitioners. If present, an adverse effect of feedback on referral rates should be particularly evident for the good responders to the feedback. With good responders we mean those general practitioners who had the strongest change in test ordering behaviour according to the recommendations in the feedback. The change in test ordering behaviour was assessed in the following way.

In the feedback provided since 1985, 11 tests were frequently and repeatedly designated as unnecessary or inappropriate for specific indications. The response of each general practitioner to the feedback was assessed by determining the percentage of decrease in the number of requests for these tests (1987 compared to 1984). Based on this percentage, general practitioners were split up into 4 quartiles. The referral patterns of the upper (good responders) and the lower quartile (poor responders) were compared.

Statistical analysis

The correlation between test ordering behaviour and referral behaviour in every year was determined by means of the Spearman rank correlation coefficient. Whenever the relation between a change in test ordering behaviour and a change

in referral behaviour was studied, trends were compared. Trends per general practitioner were taken to be the average change per year as calculated by linear regression of numbers of requests and referrals versus year.¹⁴ Significance of the differences in trends per general practitioner between individual specialisms was tested by means of the Wilcoxon signed ranks test. Significance of the differences in trends between good and poor responders was tested by means of the Mann-Whitney test. A two-sided significance level of 0.05 was used in all calculations.

Results

Of all 85 general practitioners affiliated to the DCC, 82 agreed to the analysis of referral data. Several general practitioners retired from practice or started their practice during the study. As a result they were not working during the whole study period. Therefore their data were excluded. Finally, 64 general practitioners were included in the study.

We compared test request numbers from the 64 general practitioners with their referral numbers over the period 1986-1991. The correlation is clear (table 1).

Table 1. Correlation between the test ordering behaviour and referral behaviour of 64 general practitioners; 1986-1991

Year	Correlation (r)	Significance (p)
1986	0.74	< 0.001
1987	0.65	< 0.001
1988	0.69	< 0.001
1989	0.57	< 0.001
1990	0.50	< 0.001
1991	0.56	< 0.001

The strongest correlation was found in 1986 (figure 1). The scatterplot shows a relatively linear correlation. Apparently, also in our region the big spenders on test use are also those with the highest referral rates.

The possibility of a higher referral rate due to a lower use of diagnostic tests was studied next. Together with the decrease in the use of tests, the number of first referrals appears to decrease as well (regression coefficient:0.68, SE:0.13, $p < 0.001$).

Figure 1. Test ordering behaviour compared with referral behaviour for the 64 general practitioners in 1986

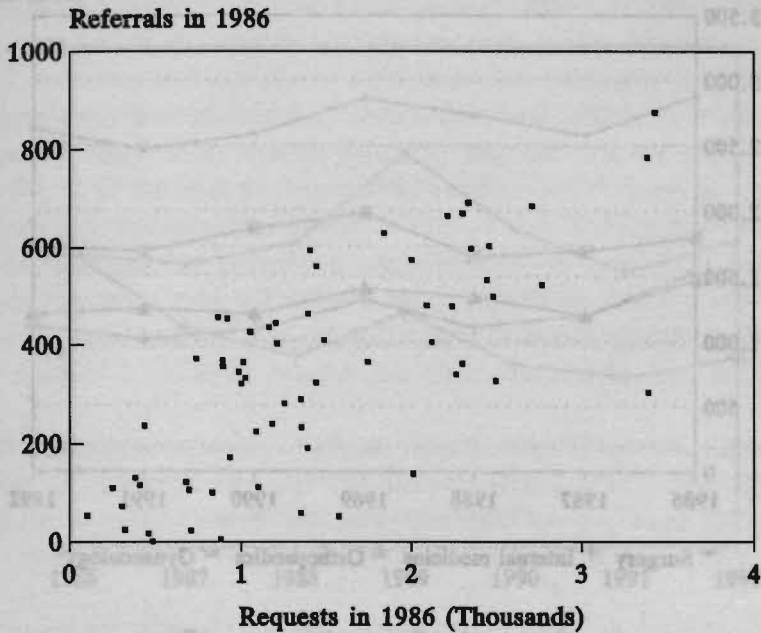


Figure 2 shows the trends in referrals for internal medicine, surgery, orthopaedics and gynaecology. Although the absolute numbers of referrals are quite different, trends are not ($p>0.06$).

When the referral trends of good responders ($N=16$) is compared with that of poor responders ($N=16$), a small but significant difference is seen (figure 3). Typically, in contrast to our basic assumption, the physicians who responded best to the feedback show a decrease in the number of referrals, whereas those who responded poorly had a slightly increasing number of referrals. The difference in trends between both group was significant ($p=0.01$).

Figure 2. The trends in first referrals by the 64 general practitioners to four selected specialisms

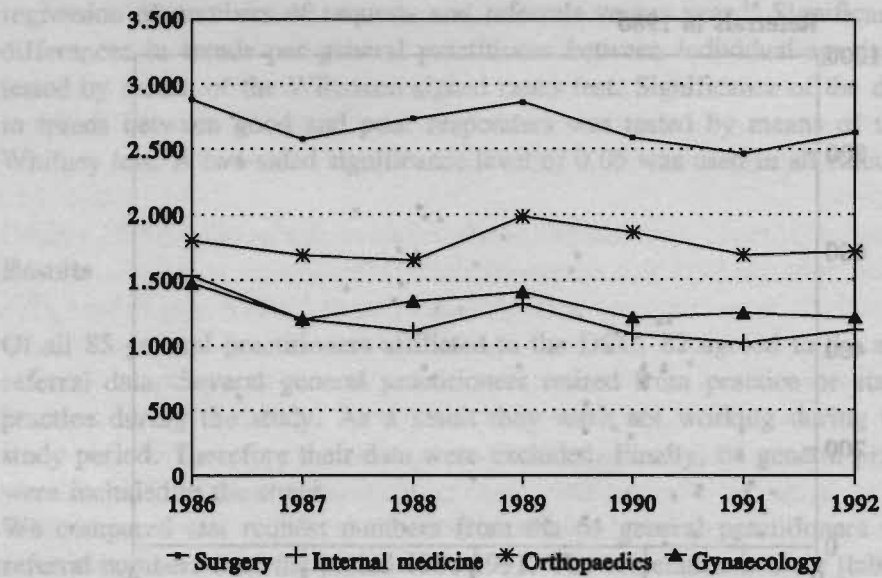


Figure 3. Trends in first referrals. Good responders to feedback (N=16) versus poor responders (N=16, indexed to the level of good responders in 1986)

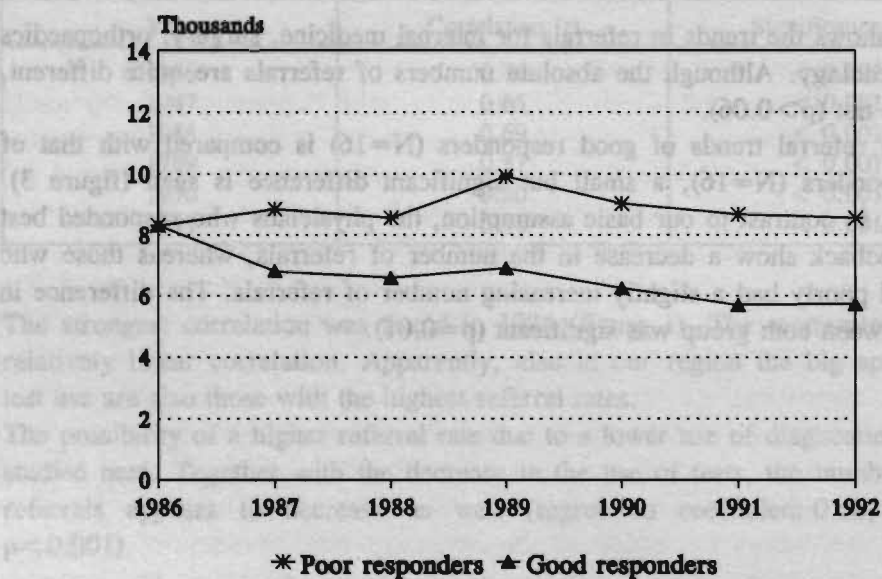
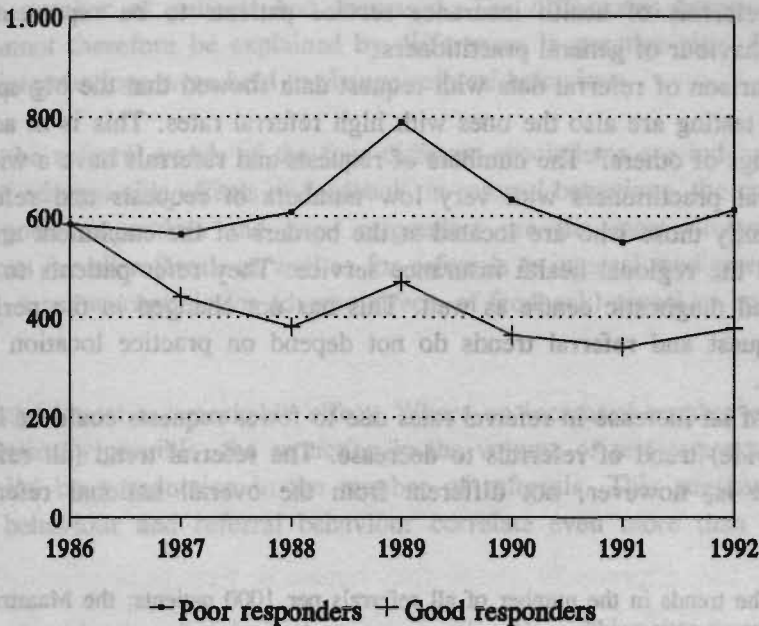


Figure 4. Trends in first referrals to internal medicine. Good responders to feedback (N=16) versus poor responders (N=16, indexed to the level of good responders in 1986)



We analyzed the referral trends to internists separately (figure 4). In the years that the feedback was provided, a small decrease was seen in the number of referrals by the good responders. In the same period, however, the number of referrals increased slightly for the poor responders. The difference between these referral trends of good and poor responders was significant ($p=0.01$).

Discussion

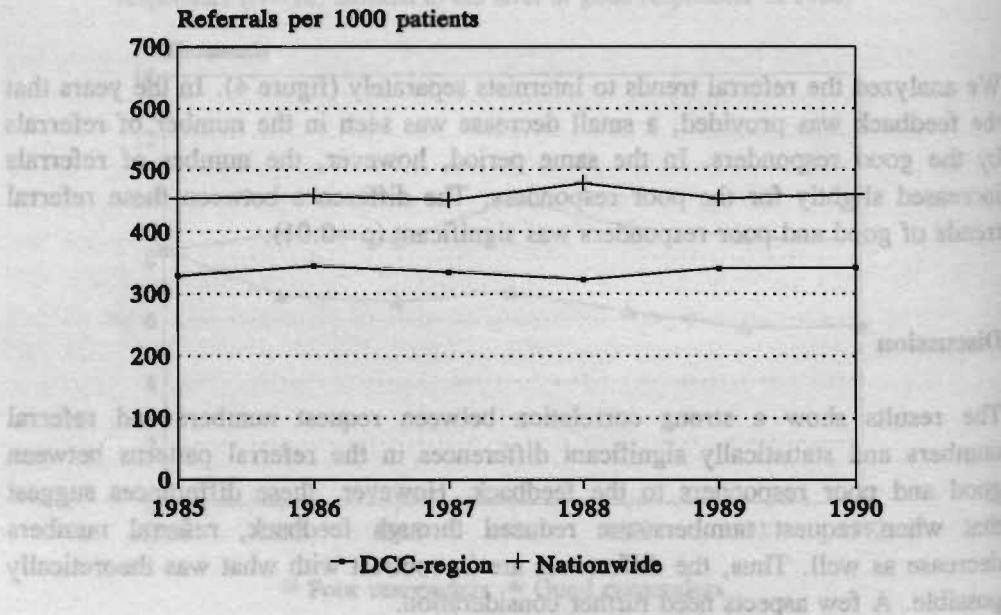
The results show a strong correlation between request numbers and referral numbers and statistically significant differences in the referral patterns between good and poor responders to the feedback. However, these differences suggest that when request numbers are reduced through feedback, referral numbers decrease as well. Thus, the differences are in contrast with what was theoretically possible. A few aspects need further consideration.

Health insurance service data do not automatically reflect the whole population. In the Netherlands 60% of the population is compulsorily insured by health insurance service. The remaining 40% is insured by private insurance companies. We consider referrals of health insurance service patients to be representative for referral behaviour of general practitioners.

Our comparison of referral data with request data showed that the big spenders in diagnostic testing are also the ones with high referral rates. This is in accordance with findings of others.⁶ The numbers of requests and referrals have a wide range. The general practitioners with very low numbers of requests and referrals are predominantly those who are located at the borders of the catchment area of our centre and the regional health insurance service. They refer patients to a nearby hospital and diagnostic centre as well. This has not changed in the period 1986-1991. Request and referral trends do not depend on practice location (data not presented).

Evidence of an increase in referral rates due to fewer requests could be hidden by a (nationwide) trend of referrals to decrease. The referral trend (all referrals) in our region is, however, not different from the overall national referral trend (figure 5).

Figure 5. The trends in the number of all referrals per 1000 patients; the Maastricht region versus nationwide



A change in test use and referral rate could be caused by a change in practice size. From our data, we had no indication whatever that in the group of good responders the number of patients had decreased while in the poor responders group the number of patients had increased. The differences between the two groups cannot therefore be explained by differences in practice size. During our study no interventions were held to change referral behaviour.

Although the referral trends of the four different specialisms are indicative of the absence of adverse side effects of feedback on referral behaviour, the comparisons between good responders and poor responders are the most illustrative. The comparisons for all referrals as well as for referrals to internal medicine (being the specialism most susceptible for adverse effects of feedback) reveal no such adverse effect.

Our findings reveal a remarkable effect. Where an increased number of referrals was theoretically possible, the reduction in the volume of test requests is in fact accompanied by a reduction in the number of referrals. This suggests that test ordering behaviour and referral behaviour correlate even more than previously expected.

Feedback should primarily aim at improving rationality of test ordering behaviour. When the advices are followed, this will not lead to more uncertainty and hence a higher referral rate. This assumption is now confirmed by the study presented in this paper. The findings cannot be applied automatically on feedback that aims solely at reducing the volume and/or expenses of diagnostic testing. It cannot be ruled out completely that an unselective reduction in the number of requests will lead to more uncertainty and thus to more referrals.

The absence of adverse effects of feedback on referral rates has important financial consequences. If referral rates had increased due to feedback, the possible financial benefits of feedback, meaning lower expenditure for diagnostic tests, would have been negated. Our data suggest that the decrease in the volume of requests through feedback is more likely to be accompanied by a lower referral rate, resulting in even higher savings in costs.

Our findings permit the following conclusions. When feedback is used to improve the quality of test ordering behaviour of general practitioners and the number of requests consequently decreases, this change is not accompanied by a higher referral rate. Whether the same holds true when feedback is provided merely to reduce the volume or costs of requests is not known.

References

1. Horder J, Bosanquet N, Stocking B. Ways of influencing the behaviour of general practitioners. *J R Coll Gen Pract* 1986; 36: 517-21.
2. Fraser CG, Woodford FP. Strategies to modify the test-requesting patterns of clinicians. *Ann Clin Biochem* 1987; 24: 223-31.
3. Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *Br Med J* 1991; 303: 398-402.
4. Buntinx F, Winkens RAG, Grol RPTM, Knottnerus JA. Influencing diagnostic and preventive performance in ambulatory care by feedback and reminders. A review. *Fam Pract* 1993; 10: 219-28.
5. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus JA. Effect of feedback on test ordering behaviour of general practitioners. *Br Med J* 1992; 304: 1093-6.
6. Schroeder SA, Myers LP, McPhee SJ, Showstack JA, Simborg DW, Chapman SA, et al. The failure of physician education as a cost containment strategy. *JAMA* 1984; 252: 225-30.
7. Kassirer JP. Our stubborn quest for diagnostic certainty. *N Engl J Med* 1989; 320: 1489-91.
8. Hartley RM, Charlton JR, Harris CM, Jarman B. Patterns of physicians' use of medical resources in ambulatory settings. *Am J Public Health* 1987; 77: 565-7.
9. Wilkin D, Smith AG. Variation in general practitioners' referral rates to consultants. *J R Coll Gen Pract* 1987; 37: 350-3.
10. Evans EO, McBride K. Hospital usage by a group practice. *J R Coll Gen Pract* 1968; 16: 294-306.
11. Mokkink HGA. Ziekenfondscijfers als parameter voor het handelen van huisartsen. Thesis, Nijmegen 1986.
12. Fertig A, Roland MO, King H, Moore T. Understanding variation in rates of referral among general practitioners: are inappropriate referrals important and would guidelines help to reduce rates? *Br Med J* 1993; 307: 1467-70.
13. Pop P, Winkens RAG. A diagnostic centre for general practitioners. *J R Coll Gen Pract* 1989; 39: 507-8.
14. Altman DG. Practical statistics for medical research. 1st ed. London: Chapman & Hill, 1991.

Routine individual feedback on diagnostic test requests - an economic evaluation -

R.A.G. Winkens¹, A.J.H.A. Ament², P.H.A. Reniers¹, R.P.T.M. Grol³, J.A. Knottnerus⁴ and P. Pop¹,

¹ Diagnostic Coordinating Center Maastricht

² Department of Health Economics, University of Limburg, Maastricht

³ Center for Research on Quality in Family Practice, Universities of Nijmegen and Maastricht

⁴ Department of Family Practice, University of Limburg, Maastricht

Submitted for publication

Abstract

Objective

To assess the economic consequences of routine individual feedback on test ordering behavior.

Data Sources

Request data were obtained from a diagnostic center and a reference laboratory. Data on costs dependent on the number of requests (variable costs) were obtained for individual tests.

Study Design

Retrospectively, we compared the cost trends in a region where feedback was provided over a 7-year period, with cost trends in situations where no feedback was provided. Also, the extra costs of the feedback were considered.

Data Collection/Extraction Methods

Differences in request trends were transformed to savings in costs of diagnostic testing.

Principal Findings

Variable costs were available for more than 90% of all requests. Since the start of the feedback in 1985, despite the extra costs of feedback, the expenditure for diagnostic testing was reduced. The savings increased as the feedback was continued. After 7 years a total net sum of more than 1.4 million US Dollars was saved in a region with 187000 inhabitants.

Conclusions

Routine individual feedback is economically worthwhile.

Key Words

Economic evaluation, feedback, diagnostic testing, quality assurance.

Introduction

The expenditures for diagnostic testing in health care are increasing year by year. Many factors, such as defensive behavior, advances in technology and the increased appeal to medical care, are responsible for this increase. Although the use of medical care and health services such as diagnostic tests increase together with health care costs, the health status of the general population in the western countries does not seem to increase accordingly.

In the last few decades several methods have been applied to improve test ordering behavior and to put a stop to the growth in the expenditure for diagnostic testing. Feedback is one of the ways to achieve these goals.¹ In several studies the effects of feedback have been demonstrated.²⁻¹⁰ In most studies the results were expressed in terms of a reduction in test use and a corresponding saving in the expenditure for diagnostic testing. Feedback apparently can influence test ordering behavior but soon after the feedback is stopped the effects diminish.^{7,10} What is also important therefore is the long-term effect of feedback.

Feedback is usually a time-consuming, and therefore expensive, method to change physicians' test ordering behavior. Based on the findings of Schroeder et al, feedback is considered not to be economically worthwhile.¹¹ The savings achieved by the reduction in test use were outweighed by the extra costs of the feedback itself. However, this conclusion need not to be representative for all situations in which feedback is provided. Feedback can be provided in various ways and thus can have different outcomes. Moreover, like many others, Schroeder provided feedback over a relatively short period of two years. If feedback is provided over a prolonged period as a routine procedure, there is great chance that the savings in costs of diagnostic testing increase and at a certain moment exceed the costs of feedback. Therefore, it is still not clear whether feedback is economically worthwhile.

The cost-effectiveness of interventions like feedback on diagnostic test requests is difficult to determine, since costs and benefits of the final outcome per patient are unknown and, moreover, depend on the underlying disorder. With regard to diagnostic testing, the underlying disorders can vary widely.

An alternative method to determine the economic consequences of feedback on diagnostic actions is to determine the expenditure for diagnostic tests before and during the feedback and to estimate the losses or savings in costs, taking into account factors of influence.

Since 1985, the Diagnostic Coordinating Center Maastricht has been providing individual feedback as a routine procedure.¹² Although the predominant goal of this feedback is to improve the rationality of test ordering behavior, it also appeared to have effects on the number of requests. Overviews showed that the start of the feedback was followed by a reduction in the total number of tests. Several tests that were discussed in the feedback frequently (and were designated in the feedback as being inappropriate for the indication for which they were requested) decreased markedly with reductions up to 95% after two years of feedback.¹³ Other tests decreased as well, probably as a result of a general learning effect. Even in the last overviews (until 1992) a persisting decrease in the number of requests was seen. These results undoubtedly had financial implications which were only indicated roughly until now.¹³ Therefore, we analyzed the economic consequences of feedback in more detail.

We addressed the following questions:

- What are the economic consequences of routine feedback?
- Is feedback worthwhile, or in other words: are the costs of the feedback itself outweighed by the savings in expenditure for diagnostic testing?

Methods

Background

The Diagnostic Coordinating Center Maastricht coordinates all diagnostic requests of the 85 affiliated family physicians serving a catchment population of currently 187000. Since 1985, every individual family physician in the area has been provided twice a year with feedback on test ordering behavior. In the feedback comments are made about inappropriate test requests and recommendations are offered for a rational diagnostic work-up. The feedback is based on request forms filled in by the family physician in the course of one month. Thus the feedback concerns real cases from daily family practice. As the request forms offer clinical data on the patient (history, physical findings, suspected diagnosis, known diseases, reason for request, etc.), it is feasible to assess the rationality of tests ordered.

The study design

Retrospectively we determined the economic effects of feedback in trend analyses on two different levels.

In a first estimation, the actual cost development in our region after the feedback was started was compared with an estimation of what the costs would have been if no feedback had been provided. This trend was estimated by using data from a comparable situation elsewhere in the Netherlands where no feedback was provided in the same period.

Second, the full extent of the economic consequences was assessed by determining the cost development for the total group of tests available at our center. In this cost development, the expenditure for the feedback was considered as well. The cost development in the Maastricht region was compared with the estimated national trend of the costs of diagnostic testing.^{14,15}

In both trend analyses several factors were considered. In the Netherlands there are fixed charges per test but it is generally assumed that these charges do not reflect the actual situation, especially when the numbers of tests change.

When this volume of tests changes, some cost factors remain relatively constant, such as manpower. Other costs vary with the numbers of tests. In order to determine the actual change in costs due to the feedback we only considered these variable costs, among which costs of materials (disposables, chemicals, etc.), equipment (by interest and depreciation). Costs of manpower were not considered. This probably leads to an underestimation of the total cost savings.

In the assessment of cost developments and savings through feedback, the expenditure for the feedback was considered. Obviously, the feedback procedure as used by the Diagnostic Coordinating Center Maastricht is not without costs. The efforts of an expert reviewer are needed and administrative support is necessary. Although only responsible for a small part of the costs, materials and postal charges are needed. All these expenses are considered.

Data collection

From the Diagnostic Coordinating Center Maastricht data were available on requests from the period 1983-1991. Comparable data were available from a laboratory elsewhere in the Netherlands, also from the period 1983-1991. This

laboratory was comparable to our center with regard to the degree of urbanization and the average family practice size. Data could be compared for 44 tests that were requested from both laboratories in the period 1983-1991. These tests were predominantly hematological, serological and clinical chemistry tests. Together they represent approximately 70% of all requests by family physicians in the Maastricht region.

Data on variable costs per test were obtained from the administrative department of the University Hospital Maastricht. These variable costs were determined at one single point of time. If present, changes in the variable costs per test could not be considered. Data on overall request trends for primary care in the Netherlands were obtained from literature.

The analysis

For each level of cost trends, the numbers of requests per test were multiplied by the variable costs per test. As a result, a total sum per test per year was obtained. To the total sum for each level, the yearly costs of the feedback procedure were added.

The savings due to the feedback were determined on every level by comparing the expenditure (including the costs of feedback) per year with the expenditure if no feedback had been provided.

Results

Variable costs per test unfortunately could not be obtained for all tests available at the Maastricht diagnostic center. The tests for which no data on variable costs could be obtained were a few radiologic tests and all endoscopic test procedures. For the 44 tests comparable with the reference laboratory, data on variable costs were obtained for all tests. In total 630 tests are available at our center, of which only about 190 tests are requested regularly. Data on variable costs were available for more than 400 tests (65%). However, these 400 tests represent more than 90% of all requests.

The costs of the feedback we provided amounted to \$ 55000 per year, for manpower (peer reviewer and secretarial assistance), equipment and materials. The costs of feedback were relatively stable during the study period.

We compared our data with those of a reference laboratory. Whereas the costs for diagnostic requests of the 44 comparable tests almost continuously increased at the reference laboratory, the corresponding costs at our center had decreased since 1985 (table 1). In 1985, the variable costs for the 44 tests decreased by 12.5% compared to the trend without feedback. In 1991, this decrease was 57%. The marked reduction in the volume of requests was accompanied by only small economic benefits. The savings were outweighed by the costs of the feedback.

At the Diagnostic Coordinating Center Maastricht far more tests can be requested than the 44 that are comparable with the reference laboratory. To look at the complete extent of the economic consequences of routine individual feedback, the complete package of all requests must be considered. The overview of the total cost development when feedback is used to change the test ordering behavior shows that the usual increase in costs per year in the Maastricht region changed into a decrease from 1985. Based on the nationwide trend of a yearly increase of 7-8% in the costs for diagnostic testing in primary care, we estimated an increase of 51% in the costs for diagnostic testing in 1991 in the Maastricht region if no feedback had been provided (table 2).¹⁵ Despite the costs of feedback, there was money saved in every year since 1985, resulting in a total return of more than 1.4 million US dollars after 7 years of feedback.

Table 1. Costs and savings in the Maastricht region for 44 tests comparable with a reference laboratory. The costs of the reference laboratory are indexed to the level of the Diagnostic Coordinating Center Maastricht (DCC) in 1984

Year	Costs DCC*	Costs reference laboratory*	Reduction *
1983	\$ 72198	\$ 76573	-
1984	\$ 79826	\$ 79826	-
1985	\$ 70753	\$ 80799	\$ 10046
1986	\$ 59808	\$ 80858	\$ 21050
1987	\$ 52157	\$ 83487	\$ 31330
1988	\$ 51664	\$ 86784	\$ 35120
1989	\$ 55766	\$ 89586	\$ 33820
1990	\$ 49544	\$ 97256	\$ 47712
1991	\$ 42234	\$ 98011	\$ 55777

* Costs of feedback not included

Table 2. Variable costs and savings for all tests except endoscopies that can be requested at the Diagnostic Coordinating Center Maastricht. Costs according trend before and after 1984 are estimated on the basis of the nationwide trend

Year	Costs DCC	Costs according to trends	Reduction	Savings* through feedback
1983	\$ 512075	\$ 494350	-	-
1984	\$ 529938	\$ 529938	-	-
1985	\$ 488081	\$ 567842	\$ 79761	\$ 24761
1986	\$ 435243	\$ 608455	\$ 173212	\$ 118212
1987	\$ 436000	\$ 651779	\$ 215779	\$ 160779
1988	\$ 450844	\$ 698269	\$ 247425	\$ 192425
1989	\$ 463240	\$ 747971	\$ 284731	\$ 229731
1990	\$ 471558	\$ 801357	\$ 329799	\$ 274799
1991	\$ 434487	\$ 858401	\$ 423914	\$ 368914

* After subtraction of costs of feedback

Discussion

The results show that the provision of feedback was accompanied by a lower expenditure for diagnostic testing, even when the extra costs necessary for providing the feedback are considered. The savings in variable costs increase when the feedback is continued. With regard to these findings a few aspects need further consideration.

The economic consequences of feedback are not easy to assess. A cost-effectiveness study is difficult to perform and an economic evaluation can show only part of the economic consequences. Also, in our comparative study we faced the problem of missing data. Variable costs could not be obtained for all tests that can be requested by family physicians.

It must be remembered that, apart from the success or failure of feedback with regard to cost containment, so far in literature costs and savings were based on charges and not on the actual cost development in which only variable costs are considered. In an earlier study, feedback appeared to reduce the costs of diagnostic testing over a period of two years, but the amount of money spent on providing the feedback outweighed the savings.¹¹ Possibly, the savings would have increased if the feedback had been continued in that study and eventually another conclusion might have been drawn.

In our calculations, no changes in manpower of laboratory staff are considered. Theoretically, a reduction in the number of requests will at a certain moment lead to a cutback in staff. Also, the variable costs were held stable in all calculations over the period 1983-1991. Undoubtedly, due to several factors such as inflation, the variable costs will have increased slightly in this 9-year period. On the other hand, due to automation, variable costs may have decreased. Altogether, we did not expect any substantial effect on the outcome of the study when costs per test were assumed to be stable.

Ideally, the study design would have been a prospective, randomized controlled trial. In our situation, for assessing the economic consequences of feedback on test ordering behavior, such a design was not possible anymore. It was at the request of the family physicians in our region that the feedback was started in 1985. Due to lack of manpower, a prospective study then could not be performed. In the meantime the feedback led to considerable changes. Further changes are much more difficult to achieve. The law of "diminishing results" is applicable here. In our region a "new" prospective study on the economic consequences of feedback on the total of all requests would have been overtaken by events and therefore it might trace only minor additional effects.

The costs of the feedback procedure turned out to be stable in the period 1985-1991. Lower costs (higher efficiency) as a result of experience in the procedure were outweighed by the fact that more tests were involved in the feedback, which makes the costs of the feedback relatively independent from the number of tests that are discussed.

As already mentioned earlier, data on variable costs were available for tests responsible for 90% of all requests. In the remaining 10% all endoscopies are included. They are the most expensive tests that can be requested at our center. The total costs in the Maastricht region therefore are likely to be higher than those shown in table 2. For the comparison with the nationwide trend, however, the missing of data on endoscopies has no consequences, since in the nationwide data, costs of endoscopies are included only to a small extent. There are only a few regions in the Netherlands (one of them being the Maastricht region) where family physicians can request endoscopies. In the other regions patients must be referred to a specialist when endoscopy is needed. Consequently, for comparing the cost trend in the Maastricht region with the nationwide cost trend endoscopies are less important.

In 1990 and 1991, a trial was performed in which feedback was provided on tests not discussed before, among which several radiologic tests. The number of requests of these radiologic tests decreased, especially in the second year of the trial. This outcome of the trial could explain the inordinate increase in the savings in 1991.

The feedback provided during the trial was merely an extension of the usual feedback and therefore did not lead to higher costs of the feedback. Low savings are seen for the 44 tests that are comparable with the reference laboratory, despite the considerable changes in the numbers of requests published earlier.¹³ This is caused by the fact that the 44 tests are predominantly low-cost tests. The determination of the economic feasibility of routine individual feedback can only be obtained by looking at the overall picture of all requests. It appears that even when the extra costs of the feedback are considered, the feedback is economically worthwhile from the beginning.

In all calculations, only variable costs were considered without possible fees for any specialist involved. These fees are in fact variable costs as well. Unfortunately the true costs of specialist involvement (e.g. for the interpretation of an X-ray) are not known. Based on nationally fixed fees, the estimated extra savings would have been considerable.

Apart from all reductions in test use and (after taking into account the extra costs of the feedback) the corresponding savings in costs, it should be emphasized that the main purpose of our feedback was an improvement in the rationality of requests and a more appropriate test use. In our opinion, feedback should always focus on quality of care, apart from the possible economic benefits.

On the basis of the population size we can give a rough estimate of what feedback might have achieved on a national level if variable costs are considered to be equal throughout the whole country. The population size in our catchment area is now 187000 which is 1.32% of the Dutch population. In the case of a nationwide provision of feedback, a total amount of more than 106 million US dollars would have been saved in the Netherlands after providing feedback for 7 years.

We conclude that routine individual feedback on test requests of family physicians is economically worthwhile. The longer the feedback is continued, the higher the savings are. This applies only when the feedback is focused on a wide range of diagnostic tests.

References

1. Eisenberg JM. Physician utilization. The state of research about physicians' practice patterns. *Med Care* 1985; 23: 461-83.
2. Schectman JM, Elinsky EG, Pawlson LG. Effect of education and feedback on thyroid function testing strategies of primary care clinicians. *Arch Intern Med* 1991; 151: 2163-6.
3. Martin AR, Wolf MA, Thibodeau LA, Dzau V, Braunwald E. A trial of two strategies to modify the test-ordering behavior of medical residents. *N Engl J Med* 1980; 303: 1330-6.
4. Fowkes FGR, Evans KT, Hartley G, Nolan DJ, Roberts CJ, Davies ER, et al. Multicentre trial of four strategies to reduce use of a radiological test. *Lancet* 1986; i: 367-70.
5. Marton KI, Tul V, Sox HC. Modifying test-ordering behavior in the outpatient medical clinic. A controlled trial of two educational interventions. *Arch Intern Med* 1985; 145: 816-21.
6. Chassin MR, McCue SM. A randomized trial of medical quality assurance. Improving physicians' use of pelvimetry. *JAMA* 1986; 256: 1012-6.
7. Eisenberg JM. An educational program to modify laboratory use by house staff. *J Med Educ* 1977; 52: 578-81.
8. Berwick DM, Coltin KL. Feedback reduces test use in a health maintenance organization. *JAMA* 1986; 255: 1450-4.
9. Everett GD, de Bois CS, Chang PF, Holets T. Effect of cost education, cost audits, and faculty chart review on the use of laboratory services. *Arch Intern Med* 1983; 143: 942-4.
10. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med* 1990; 322: 1499-504.
11. Schroeder SA, Myers LP, McPhee SJ, Showstack JA, Simborg DW, Chapman SA, et al. The failure of physician education as a cost containment strategy. *JAMA* 1984; 252: 225-30.
12. Pop P, Winkens RAG. A diagnostic centre for general practitioners: results of individual feedback on diagnostic actions. *J R Coll Gen Pract* 1989; 39: 507-8.
13. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus JA. Effect of feedback on test ordering behaviour of general practitioners. *Br Med J* 1992; 304: 1093-6.
14. Dunning AJ. Growth in number and costs. On diagnostic and therapeutic actions in health care. The Hague, 1986.
15. Financial Overview of Care. Report nr 21310. The Hague, 1990.

Fitting a routine health care activity into a randomized trial - an experiment under non-experimental conditions? -

Summary

Due to possible methodological and practical problems many researchers refrain from the use of randomized controlled trials to evaluate facilities that are already applied in routine health care. The importance and feasibility of randomized controlled trials in such a situation is shown. We performed a randomized controlled trial on the effectiveness of routine individual feedback on test ordering behavior of family physicians. While the feedback had already been applied for 4 years, the trial was started and held for 2½ years.

Studying a routine procedure that is already in action makes it difficult to find previously unexposed study subjects. After several adaptations a randomized trial is indeed possible, including well defined intervention and control groups and blinding of study subjects. Especially when the study subjects are doctors, the use of a routine procedure offers the possibility to refrain from obtaining informed consent. Thus, a contamination of the trial arms and a Hawthorne effect can be avoided.

Introduction

Clinical trials are nowadays widely accepted as the paradigm for the evaluation of the effects of therapeutic, as well as diagnostic or preventive strategies.

In health care research, however, for practical reasons it is often not possible to perform a randomized controlled trial. Frequently used alternatives then are quasi-experimental and observational studies.¹⁻³

In by far the majority of situations a trial is performed under circumstances especially created for the benefit of that trial: doctors are invited to participate, for patients inclusion criteria are defined, tests and interventions are carried out more thoroughly or procedures are used that are sometimes superfluous in daily practice, patients are monitored longer and more systematically, etcetera.

Performing a trial requires adaptations in health care to enable a correct use of the trial design. When a health care activity itself, as is or will be applied in daily practice, is studied in a trial, it is an opposite situation. Here, the trial design must be adapted to fit within health care. This is especially true when it concerns the evaluation of an activity or object that is already applied routinely and is widely accepted.

The most important adaptations of study designs to the setting of daily practice, found in literature, are the omission of randomization and an improper selection of participants.⁴⁻⁸ At times no control group whatever is used.⁷⁻⁹ Probably, the conductors of these studies expected practical problems if the study had to be performed completely according to the rules for a randomized controlled trial.

In the evaluation of routine health care activities, adaptations of routine care are not always necessary to safeguard the success of a trial. On the contrary, favorable conditions as mentioned above are not always desirable. In routine health care, it is better to evaluate the actual situation instead of the situation under optimized conditions that possibly do not reflect the actual level of care. Also, the fact that an activity is already in use can have methodological benefits. In the following we give an example of how the effectiveness of activities that are already applied in routine daily health care can be evaluated in a randomized controlled trial. We performed this trial on the effectiveness of feedback on diagnostic test ordering by family physicians. Apart from the study questions that were to be answered by the trial, the specific methodology and the feasibility of a randomized controlled trial on a routine health care activity were important. Is such a design possible for evaluating routine health care activities? When adaptations or concessions are necessary, do they interfere with the validity of the study results?

Methodology

The questions will be answered using the example of a trial on routine feedback on test ordering behavior given by the Diagnostic Coordinating Center Maastricht.

Background and objectives

The Diagnostic Coordinating Center Maastricht coordinates all diagnostic test requests of the (± 85) family physicians in Maastricht and its surrounding area. To improve the appropriateness of test use the Diagnostic Coordinating Center Maastricht gives personal feedback on test ordering behavior to every individual family physician affiliated to the center. The feedback is provided twice a year and focuses on the number and rationality of diagnostic test requests. To that end request forms submitted by each family physician in the course of one month are discussed. Discussing rationality of requests is feasible since the request forms contain routine clinical data (history, physical examination, possible diagnosis, reason for request, etc.). Long before the trial was held, this feedback procedure had been started (1985). Therefore, at the start of the trial in 1989 the feedback had become a routine activity that was applied and integrated in day-to-day health care.¹⁰

Before the feedback was initiated in 1985 there was a continuous increase in the number of requests per year. Soon after the onset of the feedback the number of requests decreased sharply as demonstrated by an observational study.¹¹ The chronological relation between the feedback and the spectacular decrease in the number of requests could not be interpreted causally. Therefore, before advising the implementation of the feedback elsewhere, a causal relation had to be demonstrated, and a randomized controlled trial was developed to that end.

The trial

The trial concerns feedback on diagnostic testing as requested by family physicians. The description and discussions about our trial will be presented in comparison with accepted principles for clinical trials.¹² In the planning of the trial adaptations of the feedback procedure were required to meet as much as possible with these principles. In order to find a balance between obeying the "rules" for a trial and not interfering too much with routine procedures we chose a design as

described below. With regard to aspects relevant for a good randomized controlled trial we describe the difficulties that may occur in routine health care and how we managed them. The following aspects will be discussed: prospective design, study subjects, randomization and interventions, data collection on effect parameters, blinding, use of placebo, sample size, statistical analysis, intention-to-treat principle and informed consent.

A prospective study design

An important benefit of an experiment is a prospective data collection. A prospective setting, preferably using randomization over the intervention and control groups, is the best way to enable the investigator to discover and eliminate or correct possible disturbances. Then it is justified to consider an effect that is achieved to be caused by the intervention under study.

In routine health care, a prospective design can be achieved simply, as long as it concerns new interventions. However, many procedures are already applied in routine health care (long) before a trial is planned or carried out. Initial effects cannot be studied anymore. In our case, if we were to perform a prospective study, this would be complicated by the fact that the feedback had already been applied for several years. Before the experiment, feedback was predominantly given on hematological, serological and clinical chemistry tests.

A potential effect had then already been achieved before the trial could be started. However, even after 4 years several tests remained that were not, or only on rare occasions, discussed in the feedback. Therefore, the feedback that was to be provided for the trial focused on these tests (electrocardiography, endoscopy, cervical smears, allergy tests and a variety of radiologic or ultrasonographic tests).

Study subjects

Preferably a trial requires the selection of subjects who are representative for the population on which the outcome of the trial should be applied. Eventually, inclusion or exclusion criteria are needed. It may be beneficial to enrol only previously unexposed study subjects.

Patient selection need not be impaired when the trial concerns routine health care. A different situation only arises when the doctor himself is the study subject, which is frequently the case in health care research, especially within the field of quality assurance.^{13,14} The selection of study subjects is then strongly restricted due to the usually small number of doctors. Frequently the participation or enrolment of a whole group in the region where the study is carried out is needed and representativeness for other regions can only be determined afterwards.

Finding unexposed study patients normally does not entail any extra difficulty when it concerns routine health care trials. When the study subjects are doctors instead of patients, previously unexposed subjects probably are more difficult to find, since the routine procedure will generally have been accepted by most physicians. In our study, the study subjects were family physicians. Family physicians had to be withdrawn only when the exposure to the trial was shorter than a year, for example due to retirement. Earlier comparisons showed that the distribution of age, sex and mean practice size of "our" group of family physicians was not different from other groups in the Netherlands. Thus, we had no reasons to expect that the group of family physicians of the diagnostic center differed substantially from all Dutch family physicians. Further selection criteria were not needed since every family physician who requests diagnostic tests regularly at our diagnostic center receives feedback.

Randomization and interventions

Random allocation of treatment is an important aspect of conducting a trial. One group receives a treatment (intervention group), the other does not (control group). Randomization allows the most reliable assessment of the effectiveness of an intervention. Also, randomization is a proper way to avoid bias. Obviously, this holds true for trials in routine health care as well. Pocock considers the use of a randomized control group preferable in all situations whenever possible.¹² With regard to the feasibility of randomization there is no clear difference between a trial on a new intervention and a trial in routine health care evaluating a routine procedure. When the study subjects are doctors, the situation resulting from the randomization may need more attention. It is possible that two doctors, one from the intervention group and one from the control group, collaborate in daily practice closely. However, in a trial where for example information is provided only to the intervention group, this would increase the risk of contamination or information leakage, threatening the outcome of the trial. In our study, family physicians on whom the intervention was applied, had to be divided at random into two groups. The tests that were to be discussed in the experimental feedback during the trial were not divided at random since several tests were closely related to each other. Therefore two clusters of tests were made. Each group of family physicians received feedback on tests from one of these two clusters. For each cluster of tests the other group of family physicians mutually served as control group (table 1).

Table 1. Trial design: tests discussed in the trial in relation to the groups of family physicians (GP) compared

Tests	GP group 1	GP group 2
Group A: Electrocardiography Endoscopy Cervical smears Allergy tests	intervention group	control group
Group B: X-rays: - chest - cervical spine - thoracic spine - lumbar spine - pelvis/hip - knee joint(s) - ankle joint(s) - sinuses Ultrasound: - liver/biliary tract - kidneys	control group	intervention group

After randomization, both groups of family physicians were comparable for practice setting, degree of urbanization, experience and gender of the family physician (table 2).

Theoretically it was possible that a family physician belonging to the control group for a specific cluster of tests could receive information about these tests if he or she discussed requests of these tests or the feedback with a colleague belonging to the intervention group involved, resulting in a smaller difference between the intervention group and the control group. However, as we have experienced since 1985 (the start of the feedback), it was not customary for family physicians to discuss the feedback reports they received with colleagues.

Table 2. Practice and physician characteristics at randomization

	Group 1	Group 2
Number of physicians	39	40
Mean experience (in years)	15	15
Standard deviation	7.5	8.5
Gender:		
* male	35	36
* female	4	4
Practice setting:		
* solo	21	23
* group practice	11	10
* health center	7	7
Urbanization:		
* city	28	25
* rural area	11	15

Data collection on effect parameters

Effects are usually assessed by analysing data that are specifically collected for the benefit of that trial. In (routine) health care research this needs not to be different but it may offer the possibility to use data that are routinely collected in daily health care. This can save time (and money). On the other hand, there is chance that data obtained from routine health care do not match with the desired data. Consequently, additional information is needed.

In our trial, we used data on tests actually requested by family physicians in daily health care. No extra efforts were needed to improve the "participation" of family physicians. Effects on the volume of test ordering behavior (number of requests per family physician) and on the rationality of test ordering behavior (number of rational and non-rational requests per family physician) were assessed by analysis of routinely collected data. Rationality was determined by comparing request data (including clinical information) with accepted guidelines or standards.

Blinding

In a trial the outcome may be distorted when both the patient and those responsible for the intervention or evaluation know what treatment or intervention is applied. This can be prevented by double blinding. For trials concerning routine health care blinding is equally important. Blinding is, however, difficult to achieve in routine health care, especially when it concerns non-drug interventions. In our trial double-blinding was practically impossible. To give feedback we needed to know to which group every family physician belonged in order to provide the

feedback according to the schedule in table 1. The theoretical alternative could have been the following: First, all requests of which the appropriateness was to be assessed, had to be anonymized and rewritten (handwriting can be recognized). Then, all tests included in the trial would be discussed for all family physicians. The reviewer giving comments through feedback would not know the outcome of the randomization. Finally, before the feedback reports are sent to the family physicians, the part on tests requested by the control group is retained. This selection would be performed by another person, not involved in the feedback. Our way to safeguard single blindness was by performing the trial without informed consent (see below). This had also the benefit that no Hawthorne effect occurred. Simply knowing that a trial was to be performed might already influence the test ordering behavior of physicians.¹⁵ Even now, after the trial had stopped, no sign whatsoever reached us that the presence of the trial was suspected. To achieve this, the new experimental feedback was appended to the routine feedback as provided since 1985. For the family physicians the trial was therefore nothing but an extension to the feedback focusing on tests not discussed before. Such extensions were not unusual for family physicians. To prevent large differences in the amount of feedback between both groups of family physicians each group of family physicians received experimental feedback on only one cluster of new tests.

Use of placebo

In trials on drug treatment a control group (untreated) frequently receives placebos. In routine health care it concerns mostly non-drug interventions. Placebo treatment is then difficult to realize. In our trial on routine feedback, placebo feedback was not possible. However, since both groups received feedback on different test-groups and because of the secrecy of the trial, a 'single blinded' placebo effect may be reasonably assumed to be associated with the mutual 'control' feedback.

Sample size

The most important factor determining sample size is the statistical requirement: how many patients are needed to detect the minimal difference or change that is relevant to be looked for. In trials in routine health care or health care research, the same principle is applicable. When doctors are the study subjects, the availability of sufficient numbers of doctors in a region is the restricting factor. In our study, the sample size was fixed at 85, being all family physicians in our region (see also statistical analysis).

Statistical analysis

Obviously, with regard to statistical testing, trials in routine health care should not be treated differently. In health care research frequently an incorrect unit of analysis is used.¹⁶ When physicians are the study subjects, they should be the unit of analysis. We randomly assigned family physicians to the two intervention groups and the corresponding control groups. This necessitated statistical analysis to see whether the potential differences are due to the feedback and independent of other differences between the compared groups. Since randomization was on physician level, statistical testing was also on physician level.

Performing a trial in routine health care can have a negative influence on the power ($1-\beta$) of the trial. Both the number of the participants (frequently doctors) that can be enrolled and the amount of data that can be obtained are relatively fixed. This may lead to situations in which the power is too low.

The intention-to-treat principle

No matter what kind of trial in no matter what situation, the intention-to-treat principle should always be kept in mind.¹⁷ Applying this principle to our trial means that all requests from family physicians allocated to one of the two groups always should be analyzed according to this allocation. In our study, switching from one cluster of feedback to another was not possible: once a cluster of feedback was allocated to a family physician, during the trial he would never receive feedback on the other cluster of tests. Stopping the "treatment", in our situation the feedback, might be considered when a family physician showed an adverse reaction to the feedback. However, a poor response to the feedback would only urge us to give even more feedback and stopping the feedback would be undesirable. An exception was made when a family physician migrated or retired from practice. After the retirement date, data were treated as missing since retirement was unlikely to be related with the feedback that was provided during the trial.

Informed consent

According to the declaration of Helsinki, investigators "should obtain the subject's freely-given informed consent" from each patient involved in a trial, unless "the doctor considers it essential not to obtain informed consent; the specific reasons for doing so should be stated in the study protocol". In that case informed consent from each doctor should be obtained. In trials in routine health care normally informed consent should be obtained as well. There are, on the other hand, situations conceivable in which it is not absolutely necessary and where it even would be unwise to obtain informed consent.

In our trial, informed consent was not obtained for two reasons. First, informed consent is not strictly required when the procedure that is to be studied is already provided or applied in daily practice and is not changed. In our study, the feedback procedure did not change, only the contents of the feedback reports were extended. In addition, no single procedure normally available was withheld due to the trial setting. Second, after informed consent the trial would no longer be secret, what, due to a Hawthorne effect and to possible contamination of the trial arms, could be hazardous to the success of the trial. Therefore, in some routine situations informed consent would preferably not be obtained in order to offer the best possibility to assess purely the effects of an intervention.

Evaluation

The above design was used by the Diagnostic Coordinating Center Maastricht during 2½ years, from October 1989 until May 1992. Apart from the outcome of the trial, the design was successful, as is shown by the following.

It was feasible to add new information to the existing feedback. The routine procedure (feedback given twice a year since 1985) was not interfered with.

After randomization, two groups were composed that appeared to have similar characteristics. We refer to table 2. The trial could be kept secret during the whole study period. We had no sign or question whatever from the family physicians indicating that the presence of the trial was suspected.

At the end of the trial it would have been simple to resume the "old" pattern of feedback that was applied before the trial. However, due to the positive results, we decided to continue the feedback on tests discussed within the scope of the trial. This "new" feedback is now integrated into the usual feedback to all family physicians.

Although the effects of the feedback could be demonstrated, these were smaller than expected. It appeared that the feedback in the past also influenced the test ordering behavior for tests that were not discussed, probably due to a general learning effect. Therefore, part of the improvement had already occurred before the trial.

Discussion

What can be learned from our experiences? Apparently, randomized controlled trials are feasible in routine health care. There are nevertheless a few considerations which should be discussed.

What are the implications of performing a trial in routine health care? Many trials, especially those on oral drug treatment, face the problem of non-compliance. Due to the fact that for (routine) diagnostic test requests all "participating" family physicians make an appeal on the diagnostic center there appeared to be no non-compliance in our trial. The same is true for our feedback: all family physicians automatically receive feedback twice a year, independent of the trial setting. There is one uncertainty: sending the feedback reports does not guarantee that they are read. In other words, does the doctor take his medicine? Although an anonymous questionnaire showed that by far the majority (90%) of the family physicians read the reports, this does not satisfactorily rule out uncertainty.

When the trial was held, it was no extra burden for the family physicians. The trial was merely an extension of what was already a routine activity. This undoubtedly reduced the possibility of non-compliance considerably. There was no specific preselection, since participation was not asked or permitted. When physicians are asked to participate, especially the most enthusiastic will agree.

In routine health care, especially when doctors are the study subjects, this selection is undesirable. Some consider this as inevitable.¹⁸ Voluntary mechanisms like continuing medical education (CME), for example, are mostly attended by those who least need it.^{14,19} A trial on CME would suffer from selection bias if physicians were asked to participate in the trial. An important benefit of implementing a trial in routine health care is the possibility of performing the trial in "secrecy", without informed consent. Some might find it unethical not to obtain informed consent. If this, however, is the best way to evaluate an activity that is already in use it might be more unethical not to perform the study. In that case, several ineffective routine activities would undeservedly be continued.

Moreover, in every prospective study, whether observational or experimental, there is the risk of a Hawthorne effect whenever a patient (or physician) is aware of the trial. This effect can lead to considerable changes which could undeservedly be attributed to the intervention.¹⁵ This risk is virtually absent when both the intervention and the control group are unaware of the presence of the trial, as in our study. On the other hand, keeping the secrecy often requires the meticulous prevention of leakage of information on the trial. This should not be underestimated.

Randomization is achievable without major problems, even when it concerns doctors. The resulting control group is the optimal group for comparison. Practical problems should be solved as far as possible to allow randomization. The possible practical problems, however, can vary widely. If a control group is taken from another region, potential confounders can sometimes not be traced and/or controlled for.

Despite the use of new items in our feedback, we experienced an influence of the routine feedback on other tests in previous years, probably caused by a general learning effect. Apparently, using the same procedure focusing on new topics that are closely related to earlier topics, is not a fully satisfactory solution. Theoretically, the best solution would have been to evaluate the feedback in a trial in another region where no feedback was provided before. For practical reasons this was not possible.

In 1985 we started the routine feedback at the request of our family physicians, not knowing what the effects would be. There are situations in which starting without having insight in the expected effects afterwards has led to unpleasant findings. PROs (Professional Review Organization) and PSROs (Professional Standards Review Organization), for example, were implemented on a large scale in the US, but failed to improve quality of care.^{20,21} Their effectiveness was insufficiently determined before the implementation.

Whenever a trial is planned on a (potential) routine health care activity, it is undoubtedly better to perform the trial before this activity is implemented.

A trial normally is held under the most favorable conditions and thus determines efficacy. Routine health care, however, optimally reflects the daily situation and trials here are useful to determine efficiency. In our opinion trials in routine health care (including care that has already been implemented) should be performed more often since many (types of) interventions are implemented but not yet evaluated properly. In our experience, it is possible to carry out such trials.

The course of our study has shown that a trial implemented into routine health care can be successful. From the trial design described above it is clear that despite some differences with the ideal trial paradigm, with a number of adaptations a randomized trial in routine health care can be performed. In fact, under routine health care conditions the results of a trial reflect the actual daily situation probably better than under artificially prepared optimized conditions.

References

1. Grady KE, Strudler-Wallston B. Research in health care settings. Beverly Hills: Sage, 1988.
2. Cook TD, Campbell DJ. Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally, 1979.
3. Bouter LM, van Dongen MCJM. Epidemiologisch onderzoek; opzet en interpretatie. Utrecht: Bohn, Scheltema & Holkema, 1988.
4. Gortmaker SL, Bickford AF, Mathewson HO, Dumbaugh K, Tirrell PC. A successful experiment to reduce unnecessary laboratory use in a community hospital. *Med Care* 1988; 26: 631-42.
5. Zaat JOM, van Eijk JThM, Bonte HA. Laboratory test form design influences test ordering by general practitioners in the Netherlands. *Med Care* 1992; 30: 189-98.
6. Anderson CM, Chambers S, Clamp M, Dunn IA, McGhee MF, Summer KR, et al. Can audit improve patient care? Effects of studying use of digoxin in general practice. *Br Med J* 1988; 297: 113-4.
7. Dyck FJ, Murphy FA, Murphy JK, Road DA, Boyd MS, Osborne E, et al. Effect of surveillance on the number of hysterectomies in the province of Saskatchewan. *N Engl J Med* 1977; 296: 1326-8.
8. Bareford D, Hayling A. *Inappropriate use of laboratory services: long term combined approach to modify request patterns.* *Br Med J* 1990; 301: 1305-7.
9. Durand-Zaleski I, Rymer JC, Roudot-Thoraval F, Revuz J, Rosa J. Reducing unnecessary laboratory use with new test request form: example of tumour markers. *Lancet* 1993; 342: 150-3.
10. Pop P, Winkens RAG. A diagnostic centre for general practitioners: results of individual feedback on diagnostic actions. *J R Coll Gen Pract* 1989; 39: 507-8.
11. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus JA. Effect of feedback on test ordering behaviour of general practitioners. *Br Med J* 1992; 304: 1093-6.
12. Pocock SJ. Clinical trials. A practical approach. Chichester: J Wiley & Sons, 1991.
13. Eisenberg JM. Physician utilization. The state of research about physicians' practice patterns. *Med Care* 1985; 23: 461-83.
14. Knottnerus JA, De Vet HCW. Effectiviteitsonderzoek in de huisartspraktijk. *Huisarts Wet* 1987; 30: 116-9.
15. Sherman H. Surveillance effects on community physician test ordering. *Med Care* 1984; 22: 80-3.
16. Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in medical care experiments. *Med Care* 1984; 22: 1101-14.

17. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol* 1992; 21: 837-41.
18. Jennett PA, Laxdal OE, Hayton RC, Klaassen DJ, Swanson RW, Wilson TW, et al. The effects of continuing medical education on family doctor performance in office practice: a randomized control study. *Med Educ* 1988; 22: 139-45.
19. Shaw D. Acceptability of audit. In: Smith R. *Audit in action*. London: BMJ publishers, 1992.
20. Soumerai SB, McLaughlin TJ, Avorn J. Quality assurance for drug prescribing. *Qual Ass Health Care* 1990; 2: 37-58.
21. Soumerai SB, Avorn J. Efficacy and cost-containment in hospital pharmacotherapy: state of the art and future directions. *Milbank Mem Fund Q* 1984; 62: 447-74.

The impact of interventions with feedback on others are also measured. This paper ends with conclusions and some recommendations for the application of the study findings as well as for future research.

The patient results

Individual feedback can be a tool to improve the "quality" of test ordering behaviour and at the same time it can reduce the continuously growing number of requests of the requests.

Soon after the start of the feedback in 1985 the number of requests decreased. For individual tests decreases up to 50% or more were seen.

In a non-random controlled trial, a statistically significant reduction in the number of requests of selected tests (selected tests in particular was found as well as a reduced number of non-rational requests with statistically significant differences for selected radiological tests (number some X-rays in particular).

At the end of the study feedback had been provided for more than 7 years. The feedback appeared to be effective throughout the whole period. But a single test frequently ordered against in the feedback showed a substantial increase. The increase in experience of the general practitioner correlated significantly with the getting used to act according to the recommendations offered in the feedback. The most experienced practitioners responding best to the feedback. While the feedback reduced the number of test requests, no compensatory higher number of referrals were seen. On the contrary, while request numbers decreased, referrals also seemed to decrease slightly as well. Routine individual feedback, repeated once a year, is economically worthwhile. The savings are higher than the costs. The longer it is continued, the higher the savings are.

General discussion

In this chapter, after an overview of the main findings, comments are given on the study as a whole, especially with regard to the methodology and the study results, including some political considerations. Since the individual study parts and their results have already been discussed in the corresponding separate chapters, the methodology and results are discussed only in a wider perspective. Striking similarities or discrepancies with findings of others are also mentioned. This chapter ends with conclusions and with recommendations for the application of the study findings as well as for future research.

The salient results

Individual feedback can be a tool to improve the "quality" of test ordering behaviour and at the same time it can reduce the continuously growing number of diagnostic test requests.

Soon after the start of the feedback in 1985 the number of requests decreased. For individual tests decreases up to 90% or more were seen.

In a randomized controlled trial, a statistically significant reduction in the number of requests of selected tests (cervical smears in particular) was found as well as a reduced number of non-rational requests with statistically significant differences for selected radiological tests (lumbar spine X-rays in particular).

At the end of the study feedback had been provided for more than 7 years. The feedback appeared to be effective throughout the whole period. Not a single test frequently advised against in the feedback showed a substantial relapse. The amount of experience of the general practitioner correlated significantly with the willingness to act according to the recommendations offered in the feedback, the most experienced practitioners responding best to the feedback. While the feedback reduced the number of test requests, no compensatory higher numbers of referrals were seen. On the contrary, while request numbers decreased, referral rates seemed to decrease slightly as well. Routine individual feedback, repeated twice a year, is economically worthwhile. The savings are higher than the costs. The longer it is continued, the higher the savings are.

The methodology

The total study consisted of 6 separate parts. From a scientific viewpoint the randomized controlled trial was the most important study part. An important feature of the feedback provided by the Diagnostic Coordinating Centre Maastricht, namely the fact that it has become a routine and widely accepted activity, may have had a negative influence on the results of the trial. Whereas the feedback dramatically changed the request patterns for tests that had been frequently discussed since 1985, the effects achieved in the trial (on a selection of the remaining diagnostic tests that were not discussed before the trial) are much less clear-cut. Partially this is probably due to a general learning effect in the years before the trial, influencing the test ordering behaviour for tests that were not discussed. Another explanation for the smaller effect of the feedback provided in the trial is that some of the tests involved in the trial were requested less often and were requested less non-rationally before the trial. Thus, advising against their use or improving the rationality of their requests was less often needed.

Altogether it seems wise to perform a randomized controlled trial before an intervention is implemented into routine care, although it proved to be possible to perform a trial afterwards, notwithstanding the inevitable practical problems. This seems obvious but the many situations where an intervention is implemented without testing its effects first demonstrate the opposite. PROs (Professional Review Organizations) and PSROs (Professional Standard Review Organizations) in the US are striking examples.

One of the factors not evaluated in the study is the effect of feedback on patient outcome.¹ In research on diagnostic testing, this aspect is extremely difficult to study. What measures should be used to detect effects of a change in test ordering behaviour on patient outcome? Also, diagnostic testing, especially in general practice, involves a broad range of clinical problems, making it very difficult to trace the final outcome for individual patients.

A vague impression of patient outcome can be obtained from the number of abnormal test results. A negative effect of feedback (and the ensuing reduction in test use) on patient outcome could be expected if the proportion between normal and abnormal test results changed in the direction of a lower percentage of abnormal test results. In that case there might be a negative selection. Data from the Maastricht diagnostic center reveal no negative influence on the percentage of abnormal test results (table 1).

Table 1. Distribution (percentage) of normal and abnormal test results and their relation in the period 1983-1991

Year	normal (%)	abnormal (%)	ratio normal/abnormal
1983	74.2	25.8	2.88
1984	74.6	25.4	2.94
1985	72.7	27.3	2.66
1986	63.2	36.8	1.72
1987	69.8	30.2	2.31
1988	71.1	28.9	2.46
1989	70.8	29.2	2.43
1990	70.9	29.1	2.44
1991	73.6	26.4	2.79

In cases where the use of a test is advised only for specific indications (for which that test has a good validity) we might expect especially a decrease in the number of normal test results. Then, the proportion of normal results would decrease whereas the proportion of abnormal results would increase.

A striking example would be Alanine Aminotransferase (ALAT) which was only recommended for several specific indications.

Table 2. Normal and abnormal test results for ALAT, their proportion and their relation in the period 1983-1991

Year	normal	abnormal	normal (%)	abnormal (%)	ratio normal/abnormal
1983	3976	858	82	18	4.63
1984	4516	656	87	13	6.88
1985	3648	552	87	13	6.61
1986	2854	446	86	14	6.40
1987	2554	714	78	22	3.58
1988	2430	730	77	23	3.33
1989	2371	601	79	21	3.94
1990	2216	784	74	26	2.83
1991	1960	734	73	27	2.67

Table 2 shows a strong decrease in the number of normal test results, whereas the number of abnormal test results remain relatively stable. Typically, this effect is achieved only after several years.

These data on test outcome must, however, be treated with caution. Eisenberg concluded that the test results have a limited value in predicting effects of changes in test use on patient outcome.² Also, in general it must be considered that of all normal or abnormal test results, a considerable number of test results can be false negative or false positive. Whether this number has changed cannot be determined from our data. Therefore, patient outcome is still a factor of uncertainty in this study.

Kroenke et al have focused on underutilization as a measure for adverse effects on patient outcome.³ In their study, no underutilization was found, but there may be doubt about the validity of the degree of underutilization as a measure for adverse effects of interventions in test use on patient outcome.

In the study no comparison was included with other strategies, such as education or financial incentives, that are also used to change the behaviour of general practitioners.

Introducing other techniques in our region would be possible, but on the other hand this could lead to disappointing results. In our region test ordering behaviour has already changed considerably in the past few years, leaving less room for any further improvement through another intervention.

Comparing the feedback study presented in this book with studies on other intervention techniques held elsewhere would be somewhat like comparing apples and oranges. The different study designs, study subjects and study settings vary widely. Therefore there is still unsatisfactory information on what type of intervention in the test ordering behaviour of general practitioners is the most effective or efficient.

In the Netherlands, there is a mixture of practice settings, however, with regional differences. All general practitioners have direct access to diagnostic testing facilities in a (hospital) laboratory or diagnostic centre. The facilities that are available for general practitioners, however, vary widely.

In the Maastricht region, the general practitioners population is also a mixture of practice settings, not very different from the situation nationwide. Thus, we considered our study to be representative for the total population of Dutch general practitioners.

Even before the feedback started, the general practitioners in the Maastricht region appeared to request fewer tests than their colleagues elsewhere. This could be explained by the guidelines distributed by the Maastricht diagnostic centre since

1979.⁴ These guidelines reflected the most rational diagnostic work-up for several clinical problems and could lead to a reduction in test use. Representativity is, however, not impaired by this lower test use prior to the feedback. Moreover, a lower test use at the baseline measurement could only cause an underestimation of the effects of the feedback. The effects of feedback elsewhere would possibly be even higher.

The study findings and the literature

In accordance with findings of others, the feedback study shows that individual feedback can reduce the number of (unnecessary) test requests. In general it is assumed that reducing the volume of test requests leads to an improvement of quality of care as well.¹ We were able to demonstrate such an improvement of rationality.

The effects on rationality we found were small, as a result of factors such as the limited choice of tests that could be discussed in the trial and the changes by the previously mentioned general learning effect prior to the trial.

A same causal relation between individual feedback and a change in test ordering behaviour was found with regard to the number of tests ordered. We consider these changes brought about by the trial to be admissible as evidence for the effects of feedback in the first years of the feedback on other tests.

The most striking effects were found with regard to the duration of the feedback. The intervention presented in this book had, by the end of the study, been in action for more than 7 years. Griner provided feedback for almost the same period but in that study the feedback was combined with a variety of other interventions.⁵ From that study the effects of the feedback therefore could not be traced.

In fact the study presented in this book is one of the first studies that confirms the persistence of effects when feedback is provided as a repeated routine procedure. As soon as the feedback is stopped, the effects are likely to diminish partially or even completely.^{6,7} Apparently it is worthwhile to continue feedback once it is started.

Despite the gain in clarity concerning the question of how to deal with inappropriate test use, the test ordering behaviour of general practitioners still has its black box. Through feedback it is only possible to look at what is actually requested appropriately or inappropriately. Unfortunately, we know little about what is not requested in situations where this should have been done. For an optimal look at test ordering behaviour, it should be possible to consider this black box as well.

Altogether we can only presume that the feedback provided by the Diagnostic Coordinating Centre Maastricht has not resulted in any underutilization of tests.

Ideally, the extent to which each individual general practitioner responds to the feedback, reflected by the drop in the number of requests, should be corrected for practice size. Otherwise the change in the absolute numbers of tests per general practitioner would give an incorrect impression of the effects of feedback. Unfortunately, the data on practice size we obtained from the questionnaire were not sufficiently reliable. Therefore, such a correction was not possible. As a solution we used the relative change (in terms of percentage) in test ordering behaviour as a measure for the response to the feedback by calculating the percentage of the decrease or increase.

In chapter 6 emphasis was put on explanatory factors. First, the relation between test ordering behaviour and practice characteristics and/or practitioner characteristics was studied. In literature, many other factors are considered to explain differences in test ordering behaviour, but if one looks at all these studies, the conclusions appear to be contradictory.⁸⁻¹² There are no practice characteristics or practitioner characteristics found in all studies that explain a high or low use of diagnostic tests in the same direction, even when at times the same questionnaire is used.

The present study also evaluated why general practitioners show different reactions to the feedback. As with test ordering behaviour itself only a few explanatory factors have been found. A good response to feedback is positively associated with the amount of experience and negatively associated with the mean duration of consultations. The more experienced general practitioners have higher numbers of requests but respond better to the feedback. This contrasts with Metcalfe, who stated that especially the younger doctors have a positive attitude towards feedback.¹³ As a result it is still not known sufficiently on whom an intervention like feedback should be aimed especially.

We found little influence of the amount of continuing medical education and/or literature reading by the general practitioner on test ordering behaviour. Our findings are in accordance with the findings of others who found little effects of these educational tools on physicians' behaviour. In only a few cases does continuing medical education have any effect.^{14,15} This once again supports the view that the use of continuing medical education must be considered on its merits.

A remarkable and unexpected finding was that the reduction in test use was accompanied by a lower number of referrals. We found no explanation for this finding. Apart from the high correlation between test ordering behaviour and referral behaviour, we expected a change (decrease) in test ordering behaviour caused by the feedback not to be followed by a change (increase) in referral behaviour. Perhaps those who change their test ordering behaviour according to the recommendations in the feedback are more willing to change also other aspects of care as they might reflect on their referral behaviour (and possibly prescription behaviour) as well. Besides, the decision analytical considerations discussed in the feedback are applicable to all aspects of care, not only for diagnostic testing but also for referring and many other aspects.

In both the request data and the referral data there is a striking temporary increase in the volume in 1989. The reason for this similarity is unknown. So far, the higher number of requests in 1989 was considered to be evidence of a moment of satiation. The increased numbers of requests as well as referrals, however, suggest a higher level of medical consumption by patients.

Looking over the publications dealing with efforts to change test ordering behaviour, mostly the effects on the volume of tests are described (see chapter 1). It appears that a reduced use and the associated savings in costs are the most important reasons for providing feedback. An improvement of the quality of test ordering behaviour should, however, be at least equally important as the reduction in the use of (mostly unnecessary) tests.

For policymakers and health care financiers, there is another important aspect, namely the saving in costs. The economic model presented in chapter 8 shows that the short-term savings are small when only variable costs are considered. Individual feedback will be economically very worthwhile when it is provided repeatedly and routinely. At a certain moment the reduction in test use will allow reductions in manpower, resulting in higher savings.

With regard to economic aspects, it should be considered that a possible positive side effect of a reduction in referrals could have further economic benefits.

Apart from this economic model there is a simple financial model, as presented briefly in chapter 5. This model shows the amount of money that was not spent and therefore has been saved in the past. For each individual test there is a fixed charge for reimbursement. Fewer tests thus lead to lower claims. Actually, these savings on the basis of reimbursements are much higher. Policymakers and health

care financiers have to decide which model, the economic or the financial, is the most appropriate. Despite possible savings in costs, the quality improvement is still the most important reason to provide feedback.

Feedback in the context of political developments

The costs of diagnostic testing are increasing each year.¹⁶⁻¹⁸ The Dutch government and several advisory boards have recognized the necessity to reduce the use of several medical facilities, including diagnostic testing.¹⁶ Research on the appropriateness and cost-effectiveness of diagnostic tests is stimulated by the Medical Developments Fund, managed by the Health Insurance Board. The Health Council aims at a more appropriate use of medical resources. It should not be automatically allowed to use any facility that is technically possible.¹⁹ In general, doctors tend to apply a new diagnostic or therapeutic facility simply because it is available.²⁰ This implies the need for a higher rationality of test ordering behaviour and a reduction of the number of unnecessary requests. The use of interventions (such as feedback) to bring about these changes is encouraged.^{18,19} The medical professional groups are advised to give guidelines or protocols and to create possibilities for peer review.¹⁸ Routine individual feedback on the basis of guidelines, as described in this book, can be one of the tools to achieve this.¹⁶ One step forward in the direction of a nationwide improvement of the quality and quantity of test requests in general practice are the standards as currently developed by the Dutch College of General Practitioners, the NHG.²¹

Conclusions

In conclusion, individual feedback can improve the quality of test ordering behaviour and reduce the number of unnecessary requests. When the feedback is provided routinely, the effects persist. The willingness to respond to the feedback increases with a higher amount of experience, the reduction in test use being associated with a (small) reduction in the number of referrals. Feedback is not only worthwhile within the scope of quality assurance, but also from an economic point of view.

Recommendations for implementation

The feedback procedure, as applied by the Diagnostic Coordinating Centre Maastricht, can be used on a larger scale. The following requirements and considerations are important.

As already proposed by others, a personal approach to every individual general practitioner is essential.²² Feedback on group level is less effective. In the provision with feedback no exceptions should be made (unless for the benefit of scientific research) and all physicians from the target group should receive feedback. Based on the findings of the study presented in this book consideration should be given to aiming the feedback more at young physicians with less experience. A decision analytical approach can reduce possible feelings of uncertainty. This may increase the response of those who prefer to take only few risks.

It is frequently stated that the person who is to give the feedback should be a respected peer.^{1,22} This peer should enjoy the confidence of the target group first, before starting the provision with feedback. Although the study did not evaluate the effect of the type of reviewer, the results are in accordance with these proposals in literature.

Obviously, for providing feedback knowledge of diagnostic testing and decision analytical principles together with sufficient clinical experience is needed.

Feedback should be based as much as possible on accepted standards or guidelines. This increases the response to the feedback. Whenever applicable, the Diagnostic Coordinating Centre Maastricht has based its feedback on regional guidelines and standards of the NHG, the Dutch College of General Practitioners. The most important advice concerns the duration of the feedback. It is clear from this study that routine feedback leads to persistent changes and that the economic return increases when the feedback is continued. Routine provision with feedback should therefore be encouraged strongly.

For feedback several facilities are required. Determination of the rationality or appropriateness of requests is only possible adequately when sufficient clinical data are provided with each request. Therefore, ideally, request forms should offer the possibility to give such clinical data. Feedback is a laborious activity. The part-time manpower of the reviewer (preferably a specialist in internal medicine) is needed, together with part-time secretarial assistance. Although not strictly necessary to enable the implementation of the feedback procedure, the formation of state-run diagnostic centres from which the feedback is provided, should be encouraged. The value of diagnostic centres is generally recognized.¹⁸

Recommendations for future research

The study presented in this book has found an answer to several questions but automatically new questions are raised that need further research.

First of all, it is clear that there is need for comparative studies, in order to find out what type of intervention in the behaviour of doctors is the one with the highest efficacy or the best efficiency. Also, despite the clear evidence for the effects of individual feedback, it is wise to search for new and potentially better tools for quality assurance.

The growing number of computers that are used in general practice offer many possibilities for new interventions in the future. When diagnostic tests, for example, in the future are requested through electronic mail using a personal computer, this computer is the ideal instrument to use supplementary interventions like reminders. Reminders are considered to be even more powerful in improving the behaviour of physicians.²³

Routine feedback now appears to be effective for at least 7 years. The request trends need to be determined for more years to follow the persistence of the effects of feedback in the future. More research is needed with regard to practice and practitioner characteristics. A great part of the discrepancies and contradictions in the findings of several studies on the influence of practice and practitioner characteristics can be explained by differences in study designs and instruments to monitor these characteristics. More uniformity in future studies is desired.

Also, the relation between (changing) test ordering behaviour and (the consequences for) referral behaviour needs more attention. An interesting, yet unsolved question is why general practitioners tend to refer less when test use is reduced through feedback.

A question that is still difficult to answer is the effect of an intervention like feedback on patient outcome. Is a patient better off when the rationality of test requests is improved or the number of unnecessary requests is reduced? Appropriate measures for patient outcome need to be found.

References

1. Eisenberg JM. Physician utilization. The state of research about physicians' practice patterns. *Med Care* 1985; 23: 461-83.
2. Eisenberg JM, Williams SV. Limited usefulness of the proportion of tests with normal results in review of diagnostic services utilization. *Clin Chem* 1983; 29: 2111-3.

3. Kroenke K, Hanley JF, Copley JB, Matthews JI, Davis CE, Foulks CJ, et al. Improving house staff ordering of three common laboratory tests. Reductions in test ordering need not result in underutilization. *Med Care* 1987; 25: 928-35.
4. Beusmans GHMI, Does RJMM. Werkafspraken in het Diagnostisch Centrum Maastricht. *Huisarts Wet* 1988; 31: 154-8.
5. Griner PF. Use of laboratory tests in a teaching hospital: long term trends. Reductions in use and relative costs. *Ann Int Med* 1979; 90: 243-8.
6. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians on the charges for outpatient diagnostic testing. *N Engl J Med* 1990; 322: 1499-504.
7. Ray WA, Schaffner W, Federspiel CF. Persistence of improvement in antibiotic prescribing in office practice. *JAMA* 1985; 253: 1774-6.
8. Forbes GB. The use of hospital laboratory facilities by general practitioners. *J R Coll Gen Pract* 1966; 11: 143-9.
9. Hartley RM, Charlton JR, Harris CM, Jarman B. Patterns of physicians' use of medical resources in ambulatory settings. *Am J Public Health* 1987; 77: 565-7.
10. Zaat JOM, van Eijk JThM. General practitioners' uncertainty and risk preference, and their use of laboratory tests. *Med Care* 1992; 30: 846-54.
11. Davis PB, Lay Yee R. Patterns of care and professional decision making in a New Zealand general practice sample. *N Z Med J* 1990; 103: 309-12.
12. Nightingale SD. Risk preference and laboratory use. *Med Decis Making* 1987; 7: 168-73.
13. Metcalfe DHH. Medical audit. *Br Med J* 1974; iii: 327
14. Strickland-Hodge B, Jeqson MH. Usage of information sources by general practitioners. *J R Soc Med* 1980; 73: 857-62.
15. Evans CE, Haynes RB, Birkett NJ, Gilbert JR, Taylor DW, Sackett DL, et al. Does a mailed continuing education program improve physician performance? *JAMA* 1986; 255: 501-4.
16. Dunning AJ. Groei in tal en last. Over diagnostische en therapeutische verrichtingen in de gezondheidszorg. Den Haag, 1986.
17. Financieel Overzicht Zorg. 21310. Den Haag, 1990
18. Biesheuvel BW. Gedeelde zorg: betere zorg. Rapport en achtergrondstudies van de commissie modernisering curatieve zorg. Den Haag, 1994.
19. Dunning AJ. Kiezen en delen. Advies in hoofdzaak van de commissie keuzen in de zorg. Den Haag, 1992.
20. Mandell HN. Technological imperative. Or, when your tool is a hammer, everything looks like a nail. *Postgrad Med* 1983; 74: 25-6.
21. Rutten GEHM, Thomas S. NHG-Standaarden voor de huisarts. Utrecht: Bunge, 1993.

22. Horder J, Bosanquet N, Stocking B. Ways of influencing the behaviour of general practitioners. *J R Coll Gen Pract* 1986; 36: 517-21.
23. Buntinx F, Winkens RAG, Grol RPTM, Knottnerus JA. Influencing diagnostic and preventive performance in ambulatory care by feedback and reminders. A review. *Fam Pract* 1993; 10: 219-28.

Summary

In this book the contents and effects of routine individual feedback on the test ordering behaviour of general practitioners are described.

General introduction and study questions (chapter 1)

Context of the study (section 1)

The Diagnostic (Coordinating) Centre Maastricht was started in 1979 and coordinates all diagnostic test requests of the ± 85 general practitioners in Maastricht and its surrounding area. At the request of the general practitioners, since 1985 individual feedback on test ordering behaviour has been provided. The first potential effects of the feedback became visible within a few years. Apparently, individual feedback can be an effective method to change test ordering behaviour of doctors. This assumption was studied in further detail.

Literature on feedback (section 2)

Feedback on diagnostic testing has been studied frequently, especially in the last 10 years. It can be effective, but this is not a matter of course. Several conditions need to be fulfilled. Feedback is more effective when it is profound and when it considers the appropriateness of tests. Feedback only on the mere number of requests or on the costs of tests is less effective. It appears that a personal approach by a respected expert peer, the use of accepted guidelines as a basis for the feedback and the acceptance of the feedback procedure by the target group are beneficial to the effectiveness of feedback, although this has been insufficiently studied through comparative studies. The effects of feedback diminish soon after the feedback is stopped. Persistent effects are seen in the few studies in which feedback is repeated. Little is known about what influences the willingness of doctors to act according to the feedback. Whether feedback is economically worthwhile is not sufficiently known. Cost-effectiveness is practically impossible to determine since changes in patient outcome are difficult to assess.

The following questions were addressed:

- Are reductions in volume or improvements in rationality restricted to tests discussed in the feedback?
- Does feedback lead to a reduction in the number of tests and/or improve the quality of test ordering behaviour?
- Can routine feedback achieve a persistent change?
- What factors explain differences between general practitioners in the test ordering behaviour and the response to the feedback?
- What was the influence of literature or continuing medical education?
- Were there any side effects on referral behaviour?
- Is feedback economically worthwhile?

The study design (section 3)

The total study as presented in this book consisted of 6 study parts. The first study was a retrospective analysis of the effects of feedback in the first years as compared to the request trends in a reference laboratory. The second study was a randomized controlled trial to determine the effects on the volume and rationality of test ordering behaviour. The third study was an observational follow-up study on the long-term effects. The fourth part of the study was a cross-sectional comparison between request data and practitioner and practice characteristics obtained by a questionnaire and an interview. The fifth study part was a longitudinal comparison of request trends with referral trends to trace any adverse side effects. The sixth study part was an economic evaluation of the effects of feedback in which the savings (compared to the request trends without feedback) were compared with the costs of the feedback procedure itself.

Feedback, a description and first overviews (chapter 2)

In chapter 2 the feedback of the Diagnostic Coordinating Centre Maastricht and its contents are elaborated.

Twice a year all general practitioners who regularly request diagnostic tests at the Maastricht diagnostic centre receive a critical report with comments on the rationality and volume of their requests in a month chosen at random. Diagnostic test requests in general and requests for individual patients, mentioned with name and date of birth are discussed. Suggestions for a more appropriate test use are offered. This feedback is based on the request forms completed by the general practitioner with each request. The request forms also contain clinical data on the

patient (history, physical examination, possible diagnosis, etc.). The main goal of the feedback is to improve the rationality of test ordering behaviour.

The onset of the feedback was followed by a marked reduction in test use. Total test use decreased with 24% after 2 years of feedback. For individual tests in that same period reductions up to more than 90% were seen.

The feedback is widely accepted and appreciated. Approximately 70% of all general practitioners react on the comments provided in the feedback, either by phone or in a letter. More than 85% were satisfied with the feedback.

Results (chapters 3-8)

Effects in the first years

The change in test request trends in the Maastricht region in the first years of the feedback compared to the trend in the region of a reference laboratory is considered in chapter 3.

Request data from the period 1983-1989 were comparable for 46 tests. For these 46 tests, the trends in the number of requests were equal in 1983 and 1984. From 1985 until 1989 in the Maastricht region there was a decrease while the increase persisted in the reference region ($p < 0.001$). The difference between both regions was the strongest for tests that were frequently discussed in the feedback ($p < 0.0001$).

Several tests, such as serum urea, that were repeatedly designated in the feedback as inappropriate, were reduced almost to zero. In the Maastricht region, tests that were not discussed decreased as well, although of course to a lesser extent. Also for these tests, the difference with the reference laboratory was significant ($p < 0.001$).

The effects of a randomized controlled trial

In chapter 4 the results are presented of a randomized controlled trial in which feedback on tests not discussed before was provided to two, randomly composed groups of general practitioners.

For the number of requests, there were significant differences between the intervention group and the control group for test-group A ($p = 0.04$). For test-group B the number also decreased for the intervention group but not significantly. Per test, significant differences were seen for cervical smears ($p = 0.01$) and for X-rays of the lumbar spine ($p = 0.02$).

An improvement of rationality was considered to be present if either the number

of rational requests increased or the number of non-rational requests decreased. There was a small, but non-significant, change in rational requests for both test-groups. Non-rational requests decreased to a significant level for test-group B ($p=0.04$), for which lumbar spine X-rays were predominantly responsible ($p=0.004$).

The long-term effects

Is there a persistent effect if the feedback procedure is continued? This question is dealt with in chapter 5.

The comparison with request data from the reference laboratory was extended to the period 1983-1991. Data were now comparable for 44 tests.

In the Maastricht region the decrease in the number of requests since 1985 was still visible in 1991. In 1991 the least tests were requested. Compared to the reference laboratory the volume of requests was reduced by 61% in 1991 ($p < 0.001$). No test frequently discussed in the feedback (and therefore with a strong decrease since 1985) showed a substantial relapse. Within the group of frequently discussed tests the strongest and most persistent decrease was seen for tests for which an alternative was recommended in the feedback ($p=0.002$).

Practitioner and practice characteristics

Not all general practitioners respond to the feedback to the same extent. Also a change in test ordering behaviour can be caused by other interventions such as continuing medical education courses or literature on diagnostic testing.

Chapter 6 shows that only few characteristics could explain differences in test ordering behaviour and in the extent to which general practitioners respond to the feedback. Test ordering behaviour depends on the number of hours per week spent in practice ($p=0.01$). General practitioners with more experience were more willing to change their behaviour according to the recommendations in the feedback ($p < 0.001$). The same was true for general practitioners with a shorter (mean) duration of consultations ($p=0.03$).

No evidence of any confounding influence of continuing medical education or literature was found.

Among the responders (96%) to the interview and questionnaire there was almost complete unanimity on the satisfaction with the Maastricht diagnostic centre and its feedback.

Adverse side effects on referrals

Theoretically it is possible that fewer requests lead to more referrals. The results of the comparison of both trends in chapter 7 show that the lower use of tests was not accompanied by a higher number of referrals. The referral trend in the Maastricht region was not different from the national trend. Also, within the Maastricht region, referral trends to internal medicine and orthopaedic surgery (disciplines for which a higher number of requests due to a lower test use could be expected particularly) were not different from those to surgery and gynaecology ($p > 0.06$). The referral trends of good responders to the feedback and those of poor responders were compared. Surprisingly, the reduction in the number of requests seen for the good responders was accompanied by a lower number of referrals, whereas for poor responders to the feedback the numbers of referrals increased slightly ($p = 0.01$).

Economic aspects

Is feedback economically worthwhile? Are the savings in costs outweighed by the expenditure for the feedback? In chapter 8 the results are presented of an economic evaluation. The savings in costs of diagnostic testing were estimated by looking at variable costs (dependent on the number of tests) after correction for the expenditure for the feedback procedure.

For the total package of all tests requested at the Maastricht diagnostic centre there were savings in costs of tests ever since 1985. The savings are not outweighed by the expenditure for the feedback. These net savings increase every year, up to almost 380000 US dollars in 1991. A total amount of more than 1.4 million US dollars was saved over the 7-year period that was studied. Apparently, feedback is indeed economically worthwhile.

Discussion (chapters 9 and 10)

The design of the trial

In chapter 9 the methodology of the trial performed in routine health care is discussed. The many practical problems that can occur require special attention. A specific problem arises when a procedure is to be evaluated that is already being applied before the evaluative study. Routine health care can on the other hand be studied without informed consent, and consequently, the study can be performed in secrecy. Any confounding by a Hawthorne effect can then be ruled out.

The study in a wider perspective

In chapter 10 the study as a whole is considered. As in all studies on feedback on test ordering behaviour, the effects on patient outcome could not be studied sufficiently. This study did not assess in what way the feedback could have its optimal effect. Also, there was no comparison with other interventions. The need for comparative studies is not reduced by the study presented in this book.

The present study is virtually the first to assess the effects of feedback if provided over (at least) 7 years. Routine feedback has persistent effects and is economically worthwhile. The question to whom feedback should be provided with more emphasis is answered unsatisfactorily. The findings can only suggest giving special attention to younger, less experienced doctors. It is remarkable that there is much disagreement on factors that explain why the test ordering behaviour differs among general practitioners.

If provided personally and repeatedly, feedback can be recommended as a procedure to change the test ordering behaviour of general practitioners. Further research should focus on comparative studies and on explanatory factors. Nonetheless, there should be constant search for better ways to change test ordering behaviour. For assessing effects on patient outcome, appropriate measures still need to be found.

Samenvatting

Het verbeteren van het aanvragen van tests in de huisartspraktijk

In dit boek worden de inhoud en effecten beschreven van routinematig verstrekte persoonlijke feedback over het aanvraaggedrag van diagnostische tests door huisartsen.

Algemene introductie en vraagstelling (hoofdstuk 1)

De context van de studie (paragraaf 1)

Het Diagnostisch (Coördinerend) Centrum Maastricht werd opgericht in 1979 en coördineert alle diagnostische aanvragen van de ± 85 huisartsen in Maastricht en omstreken. Op verzoek van de huisartsen werd in 1985 gestart met het halfjaarlijks verstrekken van individuele feedback over het aanvragen van diagnostische tests.

De eerste effecten van de feedback werden enkele jaren later zichtbaar. Kennelijk kan persoonlijke feedback een effectief middel zijn om het aanvraaggedrag van artsen te veranderen. Deze veronderstelling is in het onderzoek nader bestudeerd.

Literatuur over feedback (paragraaf 2)

Feedback op het diagnostisch handelen is vooral in de afgelopen 10 jaar vaak het onderwerp van studie geweest. Feedback kan een effectief middel zijn, doch succes is lang niet altijd verzekerd. Aan diverse voorwaarden moet worden voldaan. Feedback is effectiever naarmate deze meer diepgaand de doelmatigheid van diagnostische aanvragen bespreekt. Feedback louter en alleen over het aantal aanvragen of over de kosten van tests is minder effectief. Kennelijk zijn een persoonlijke benadering door een ervaren collega, de toepassing van geaccrediteerde richtlijnen en standaarden als basis voor de feedback alsmede de acceptatie van de feedback door de ontvangende doelgroep van groot belang voor de effectiviteit van de feedback. Deze veronderstellingen zijn echter onvoldoende getoetst via vergelijkend onderzoek.

De effecten van de feedback verdwijnen doorgaans zodra het verstrekken van de feedback wordt gestaakt. Aanhoudende effecten worden slechts gezien in de enkele studies waarin de feedback over een langere periode wordt herhaald. Er is weinig bekend over factoren die de bereidheid van artsen om de adviezen op te volgen beïnvloeden.

Of feedback economisch gezien rendabel is, is eveneens onvoldoende bekend. De kosten-effectiviteit van feedback is vrijwel niet te bepalen aangezien veranderingen in de patient outcome moeilijk vast te stellen zijn.

De volgende vraagstellingen werden bestudeerd:

- Beperken afnames in de aantallen of verbeteringen in de rationaliteit van aanvragen zich alleen tot tests die in de feedback zijn besproken?
- Leidt feedback tot een afname van het aantal testaanvragen en/of een verbetering van de rationaliteit van het aanvraaggedrag?
- Leidt routinematig verstrekte feedback tot een aanhoudende verandering?
- Welke factoren verklaren verschillen tussen huisartsen in aanvraaggedrag en opvolggedrag?
- Wat was de (verstorende) invloed van literatuur of nascholing?
- Heeft individuele feedback op het aanvraaggedrag nadelige effecten op het verwijsgedrag?
- Is feedback, economisch gezien, rendabel?

Het onderzoeksdesign (paragraaf 3)

Het volledige onderzoek zoals beschreven in dit boek bestaat uit 6 afzonderlijke deelstudies. De eerste deelstudie was een retrospectieve analyse van de effecten in de eerste jaren van de feedback via een vergelijking met de aanvraagtrends in een referentielaboratorium. De tweede deelstudie was een gerandomiseerd experiment dat een causaal verband tussen de feedback en een verandering in aantal en rationaliteit van diagnostische aanvragen moest aantonen. De derde deelstudie was een observationeel follow-up onderzoek naar de lange-termijn effecten van feedback. De vierde deelstudie was een vergelijking tussen aanvraaggegevens en gegevens over huisarts- en huisartspraktijkenmerken zoals verkregen via een interview en enquête.

De vijfde deelstudie was een longitudinale vergelijking van aanvraagtrends met verwijstrends om enig nadelig effect op het verwijsgedrag op te sporen. De zesde deelstudie was een economische evaluatie van de effecten van de feedback waarin de besparing (ten opzichte van de aanvraagtrends zonder feedback) in variabele kosten werd vergeleken met de kosten van de feedback zelf.

Een beschrijving van de feedback en eerste overzichten (hoofdstuk 2)

In hoofdstuk 2 worden zowel de manier van feedback geven zoals toegepast door het Diagnostisch Coördinerend Centrum Maastricht als de inhoud van de feedbackrapporten beschreven.

Halfjaarlijks ontvangt elke huisarts die regelmatig diagnostische tests aanvraagt via het centrum een kritisch rapport waarin het aantal en de rationaliteit wordt besproken van de aanvragen uit een recente willekeurige maand. Het voornaamste doel van de feedback is het verbeteren van de kwaliteit van het aanvraaggedrag. Aanvragen in het algemeen alsmede met betrekking tot individuele patiënten, vermeld met naam en geboortedatum, worden besproken. Suggesties voor een doelmatiger testgebruik worden aangereikt. De feedback is gebaseerd op de aanvraagformulieren die door de huisarts worden ingevuld en worden voorzien van klinische gegevens (ondermeer uit anamnese, lichamelijk onderzoek en de mogelijke diagnose) over de betreffende patiënt. De start van de feedback ging gepaard met een aanzienlijke verandering van het aantal aanvragen. Het totale aantal daalde met 24% na twee jaar feedback verstrekken. Voor afzonderlijke tests werden reducties tot meer dan 90% gezien. De feedback wordt alom gewaardeerd. Ongeveer 70% van de ontvangers reageert op de rapporten, hetzij telefonisch, hetzij per brief. Meer dan 85% gaf te kennen de feedback op prijs te stellen.

Resultaten (hoofdstukken 3-8)

Effecten in de eerste jaren

De verandering van de aanvraagrends in de eerste jaren van de feedback is beschreven in hoofdstuk 3.

Aanvraaggegevens over de jaren 1983 tot en met 1989 waren vergelijkbaar voor 46 tests. Voor deze 46 tests waren de trends in 1983 en 1984 gelijk. Vanaf 1985 tot in 1989 was er in de regio Maastricht een daling terwijl de trend van 1983 en 1984 (toename) zich gestaag voortzette in de regio van het referentielaboratorium ($p < 0,001$). De verschillen tussen beide regio's waren het grootst voor die tests welke frekwent in de feedback waren besproken ($p < 0,001$). Van enkele tests zoals serum ureum, die in de feedback veelvuldig als overbodig werden aangeduid, daalde het aantal aanvragen bijna tot 0. In de regio Maastricht daalde eveneens, zij het in mindere mate, het aantal aanvragen van tests die niet in de feedback aan de orde zijn gekomen. Ook voor deze tests waren de verschillen met het referentielaboratorium significant ($p < 0,001$).

De effecten van een gerandomiseerd experiment

In hoofdstuk 4 zijn de resultaten beschreven van een experiment waarbij feedback is gegeven aan twee gerandomiseerd samengestelde groepen huisartsen over een aantal niet eerder besproken tests.

Met betrekking tot het aantal aanvragen waren er significante verschillen tussen de interventiegroep en de controlegroep voor testgroep A ($p=0,04$). Voor testgroep B nam het aantal aanvragen eveneens af voor de interventiegroep doch het verschil met de controlegroep was niet significant. Per afzonderlijke test werden significante verschillen gezien voor cervix-uitstrijkjes ($p=0,01$) en röntgenfoto's van de lumbale wervelkolom ($p=0,02$).

Een verbetering van de rationaliteit werd aanwezig geacht indien het aantal als rationeel beoordeelde aanvragen toenam en/of het aantal als niet-rationeel beoordeelde aanvragen afnam. Er was een kleine doch niet-significante toename van het aantal rationele aanvragen. Het aantal niet-rationele aanvragen daalde wel significant en wel voor testgroep B ($p=0,04$), waarvoor vooral de aanvragen van lumbale wervelkolomfoto's verantwoordelijk waren ($p=0,004$).

De lange-termijneffecten

Houdt een effect aan als de feedbackprocedure wordt gecontinueerd? Deze onderzoeksvraag wordt behandeld in hoofdstuk 5.

De vergelijking van de aanvraaggegevens van het Diagnostisch Coördinerend Centrum Maastricht met die van het referentielaboratorium werd uitgebreid naar de periode 1983 tot en met 1991. De gegevens van 44 tests konden worden vergeleken.

In de regio Maastricht was de trend die in 1985 inzette (afname) in 1991 nog steeds zichtbaar. In 1991 werd het laagste aantal aanvragen gemeten. Ten opzichte van de trend in het referentielaboratorium was het aantal aanvragen in 1991 afgenomen met 61% ($p<0,001$). Geen van de tests die veelvuldig in de feedback aan de orde zijn geweest (en derhalve de sterkste daling na 1984 lieten zien) vertoonde een terugval naar de oude trend.

Binnen de groep van tests die veelvuldig in de feedback zijn besproken was de sterkste en langdurigste afname zichtbaar voor tests waarvoor in de feedback een alternatief werd aanbevolen ($p=0,002$).

Huisarts- en praktijkkenmerken

De mate waarin huisartsen gevolg geven aan de adviezen in de feedback verschilt onderling. Veranderingen in het aanvraaggedrag kunnen ook worden veroorzaakt door andere interventiemethoden zoals literatuur of nascholing.

Hoofdstuk 6 laat zien dat slechts weinig huisarts- of praktijkkenmerken verklaren waarom er verschillen zijn onder huisartsen met betrekking tot het aanvraaggedrag en het opvolggedrag van de feedback.

Het aanvraaggedrag op zich blijkt afhankelijk te zijn van het aantal uren praktijkwerk per week ($p=0,01$). Het opvolggedrag hangt af van de hoeveelheid (jaren) ervaring van de huisarts. Ervaren huisartsen volgen de feedback beter op ($p<0,001$). Daarnaast bleek hoe korter de gemiddelde duur van consulten was, hoe beter de huisarts de adviezen opvolgde ($p=0,03$).

Voor enige vorm van confounding als gevolg van nascholingscursussen en/of literatuur werden geen aanwijzingen gevonden.

Overigens waren de respondenten (respons 96%) vrijwel unaniem tevreden over het Diagnostisch Coördinerend Centrum Maastricht en de van daaruit verstrekte feedback.

Nadelige effecten op het verwijsgedrag

Theoretisch bestaat de kans dat het minder aanvragen van diagnostische tests leidt tot meer verwijzingen. De resultaten van een vergelijking van aanvraagstrends en verwijstrends, zoals weergegeven in hoofdstuk 7, laten zien dat het minder aanvragen van tests niet leidt tot meer verwijzingen. De verwijstrend in de regio Maastricht verschilde niet van de landelijke trend. Binnen de regio Maastricht bleken de trends in eerste verwijzingen naar interne geneeskunde en orthopedie (specialismen waarvoor een toename van het aantal verwijzingen door minder aanvragen het eerst te verwachten was) niet te verschillen van die voor chirurgie en gynaecologie. ($p>0,06$).

De verwijstrends van goede en slechte opvolgers van de feedback werden onderling vergeleken. Merkwaardigerwijs ging de afname van het aantal aanvragen van diagnostische tests voor de goede opvolgers gepaard met een lager aantal verwijzingen, terwijl het aantal verwijzingen voor de slechte opvolgers licht toenam ($p=0,01$).

Economische aspecten

Is feedback economisch gezien de moeite waard? Worden de besparingen tenietgedaan door de kosten van de feedback? In hoofdstuk 8 worden de resultaten gepresenteerd van een economische evaluatie. De werkelijke kostenbesparingen in de diagnostische aanvragen werden bepaald aan de hand van de reductie in variabele kosten (deze zijn afhankelijk van het aantal verrichte tests) na correctie voor de uitgaven voor de feedback.

Voor het totaal van alle tests die kunnen worden aangevraagd in de regio Maastricht was er met ingang van 1985 in elk jaar een besparing. Deze besparing was altijd hoger dan de uitgaven die voor de feedback nodig waren. De netto "winst" steeg jaarlijks, tot zo'n f 670.000,- in 1991. In totaal werd in het tijdsbestek van 7 jaren waarover de effecten van de feedback werden bestudeerd minstens 2,5 miljoen gulden bespaard. Kennelijk is feedback inderdaad (ook) in economisch opzicht de moeite waard.

Discussie (hoofdstukken 9 en 10)

Het design van de trial

In hoofdstuk 9 wordt de methodologie besproken van het gerandomiseerde experiment dat is uitgevoerd in de dagelijkse gezondheidszorg. Het groot aantal praktische problemen dat zich kan voordoen vereist speciale aandacht. Een specifiek probleem doet zich voor als een procedure wordt bestudeerd welke reeds is toegepast vóór het onderzoek plaatsvindt. In de dagelijkse gezondheidszorg kan echter onderzoek mogelijk worden uitgevoerd zonder informed consent. Dientengevolge is het mogelijk onderzoek uit te voeren zonder dat de onderzoekspopulatie ervan op de hoogte is. Op die manier kan een Hawthorne-effect worden voorkomen.

Het onderzoek in een breder perspectief

In hoofdstuk 10 wordt het onderzoek in zijn algemeenheid besproken. Zoals in alle studies op het gebied van feedback ter verandering van het aanvraaggedrag konden de effecten op de patient outcome niet voldoende worden bestudeerd. In het gepresenteerde onderzoek kon eveneens niet worden nagegaan wat de beste manier van feedback verstrekken is. Tevens was er geen vergelijking met andere interventiemethoden. Het gebrek aan vergelijkende studies is dus niet kleiner geworden door het beschreven onderzoek.

Het onderzoek was nagenoeg het eerste dat de lange-termijn effecten over een periode van 7 jaren bestudeerde. Routinematig verstrekte feedback heeft langdurige effecten en is economisch rendabel. De vraag aan wie feedback vooral moet worden gegeven kon onvoldoende worden beantwoord. De bevindingen wijzen alleen op de noodzaak om feedback vooral aan minder ervaren huisartsen te geven.

Daarnaast is het opvallend hoe tegenstrijdig de conclusies zijn van onderzoek naar factoren die de verschillen in aanvraaggedrag onder huisartsen kunnen verklaren.

Dankwoord

Hoewel het schrijven van een proefschrift doorgaans een nogal solistische bezigheid is, wordt de schrijver gelukkig door een aantal mensen met raad en daad bijgestaan. Velen leverden bovendien een bijdrage aan het betreffende onderzoek. Een aantal van al deze personen wil ik noemen en bedanken.

Om te beginnen de medewerkers van het DCC. Vanaf 1986 was het DCC een prettige thuisbasis voor mijn werkzaamheden. Zonder het DCC was dit onderzoek en het proefschrift er ongetwijfeld niet geweest. Wellicht heeft dit boek voor eenieder van jullie iets herkenbaars.

In het onderzoek is veelvuldig gebruik gemaakt van gegevens van anderen.

De medewerkers van het referentielaboratorium A, in het bijzonder de directie en databeheerder, dank ik voor de uiterst waardevolle gegevens over diagnostische aanvragen door huisartsen in hun werkgebied. Ik heb beloofd nimmer namen te noemen. Helaas wordt deze dankzegging daardoor minder persoonlijk.

Gegevens zijn ook verkregen van het academisch ziekenhuis Maastricht, de Dienst Administratie en Bedrijfseconomie in het bijzonder. Het VGZ te Maastricht (ir J. Kaasenbrood) stelde verwijscijfers ter beschikking.

Dankzij de medewerking van een aantal deskundigen konden nieuwe werkafspraken worden samengesteld ten behoeve van het gerandomiseerde experiment. Mijn dank gaat daarvoor uit naar dr A. Gorgels (Cardiologie), prof. dr G. Essed (Gynaecologie), dr ir P. Menheere (Klinische Chemie), drs E. van der Linden en prof. dr J. van Engelshoven (Radiologie).

Adviezen omtrent de vergelijkbaarheid van aanvraaggegevens werden verkregen van het De Wever Ziekenhuis Heerlen (prof. dr P. Brombacher), van het voormalige Landelijke Bureau Coördinatie 1^e-2^e lijn (mw V. Wittenberg) en van de Samenwerkende Artsenlaboratoria in Nederland (SAN).

Bij het samenstellen van de vragenlijst werden adviezen verkregen van dr F. Nieman (azM) en drs K. van der Velden (NIVEL, Utrecht).

Dr A. Ament van de vakgroep Economie van de Gezondheidszorg bood ondersteuning bij de economische evaluatie.

Voor het onderzoek moesten ongeveer 200.000 aanvraagformulieren worden opgeslagen in een databestand. Vic Dreessen (vakgroep Medische Informatica, RL) creëerde via Turbo Pascal een efficiënte en betrouwbare database voor dit doel. Gitte Pfeiffer verzorgde naderhand de conversies waardoor de databestanden konden worden geanalyseerd. Zonder haar hulp hadden de analyses aanzienlijk langer geduurd.

Diana Kerner, Arja Smeets, Nicole van Wissem, Ine Willemsen, Erna van Dijk, Lia Trags en Daisy Giesen waren ijverige datatypisten. Mede dankzij hun nauwkeurigheid is een omvangrijke en betrouwbare database verkregen.

Voor de correcte afhandeling van allerlei logistieke zaken en een zorgvuldig financieel beheer van het project zorgde Peter Reniers.

Het slagen van een project is mede afhankelijk van het functioneren van een projectgroep. Tevreden kijk ik terug op de 4½ jaar dat deze groep heeft gefunctioneerd. Ellen Aelfers, George Beusmans, Annemiek Bugter, Richard Grol, Guillaume Heemels, Arnold Kester, Guido Konings, André Knottnerus, Peter Pop en Peter Reniers dank ik voor hun enthousiasme en inzet.

Het omvangrijke feedbackproject heeft drie assistent-onderzoekers "versleten". Eerst Ellen Aelfers, daarna Guido Konings en tenslotte Annemiek Bugter waren op vele fronten inzetbaar en hebben veel belangrijke hand- en spandiensten verricht.

Gerda Fijten dank ik voor het kritisch nalezen van de voorlaatste versie van het manuscript en voor de vele momenten van "even overleggen" die er in de afgelopen jaren zijn geweest. Gelukkig waren niet alleen onze projecten onderwerp van discussie.

Voor de correcties van mijn onvolkomenheden in het gebruik van de engelse taal voor dit proefschrift kon ik rekenen op de hulp van David Warndorff en Bob Wilkinson.

In de analyses werd ik bijgestaan door Arnold Kester. We zijn er samen in geslaagd om in het oerwoud van data de juiste bomen te vinden.

Het vele typewerk en de lay-out van dit proefschrift werden verzorgd door de secretaresses Simone van Steenberg en Isi Vos.

Voor de samenstelling van dit proefschrift kon ik rekenen op de steun van drie promotores. Van hen heb ik veel geleerd.

Peter Pop, jouw werk in de transmurale gezondheidszorg was de direkte aanleiding voor mijn onderzoek. Dit proefschrift is dan ook mede jouw verdienste. Op jouw onvoorwaardelijke hulp en toewijding kon ik ten allen tijde rekenen. Daarvoor mijn dank.

André Knottnerus, in jouw drukke bestaan wist je altijd nog tijd vrij te maken. Jouw kennis en ervaring, niet alleen op het gebied van onderzoek, waren zeer inspirerend.

Richard Grol, jouw inbreng op het gebied van kwaliteitsverbetering gaf een extra dimensie aan het onderzoek en verruimde mijn blik. Ik heb er veel van geleerd en hoop dat je naast een kwaliteitsverbetering van het aanvraagdgedrag ook een kwaliteitsverbetering van de onderzoeker hebt kunnen waarnemen.

Mijn familie en vrienden dank ik voor de vele aandacht en de broodnodige afleiding. In dankbaarheid gaan mijn gedachten terug naar mijn ouders; zij zouden dit boek vol trots hebben bekeken.

In een dankwoord neemt de waardering voor de inbreng en de steun van mensen vaak toe hoe verder de tekst vordert. Vandaar dat ook hier het thuisfront als laatste genoemd wordt. Anja, in een tijd dat jij meer aandacht van een echtgenoot had horen te krijgen, was het eerder andersom. Geduldig en steeds weer geïnteresseerd luisterde je als me weer iets hoog zat of data me helemaal in beslag namen. Je steun was de stimulans om door te zetten.

Curriculum Vitae

Ron Winkens was born on May 19th, 1959 in Heerlen, the Netherlands. He completed secondary school at the St. Antonius Doctor College in Kerkrade in 1977. In that same year he entered medical school at the University of Limburg. He graduated in 1983, and subsequently undertook vocational training from 1984 until 1985. Soon thereafter he attended the one-year Training in Research for General Practitioners at the Department of General Practice, finishing in 1986. Since 1986 he has been working at the Diagnostic Coordinating Centre Maastricht as a researcher and general practitioner (part-time). It was here that the study presented in this book was performed. In 1988 and 1989 he was a member of the working group preparing the standard for urinary tract infections of the Dutch College of General Practitioners (NHG). Since 1987 he has been involved in continuing medical education as a member of the regional working group on CME for general practitioners (WDH).

Voor de correctie van mijn onvolkomenheden in het gebruik van de Engelse taal voor dit profielstuk kan ik u danken op de hulp van David Warneford en Paul Wilkinson.

In de analyses werd ik bijgestaan door Arnold Kester. We zijn er samen in geslaagd om in het oerwoud van data de juiste bomen te vinden.

Appendix I

The request form of the Diagnostic Coordinating Centre Maastricht

diagnostisch centrum



Sideling D.C.C. Academisch Ziekenhuis Maastricht
Tel: 043 - 875385

☐ CTO

Lokatie patiënt-contact :

☐ aan huis

☐ bij huisarts

☐ ziekenhuis

Openingsijden 08.00-16.00 uur

Datum:

Geb.dat:

Naam:

Adres:

Woonpl:

ZF + nr:

Aanvrager:

Voort:

☐ ER

PROBLEMEN / KLACHTEN VAN DE PATIËNT	BEVINDINGEN HUISARTS BIJ LICH. ONDERZOEK
-------------------------------------	--

VERMOEDELIJKE DIAGNOSE / VRAAGSTELLING	MEDICATIE	REDEN VAN AANVRAAG
		<input type="checkbox"/> Ter bevestiging aandoening <input type="checkbox"/> Ter uitsluiting aandoening <input type="checkbox"/> Screening onduidelijke pathologie <input type="checkbox"/> Controle bekende aandoening <input type="checkbox"/> Ter geruststelling <input type="checkbox"/> Check up <input type="checkbox"/> Op verzoek patiënt <input type="checkbox"/> Op advies/verzoek specialist
REEDS BEKENDE AANDOENING(EN)		

CHEMISCH BLOEDONDERZOEK <input type="checkbox"/> Glucose (nuchter) 3.3-5.2 <input type="checkbox"/> Glucose v.m. < 8.8 <input type="checkbox"/> Glucose n.m. < 8.8 <input type="checkbox"/> Natrium 135-142 <input type="checkbox"/> Kalium 3.5-4.9 <input type="checkbox"/> Calcium 2.1-2.8 <input type="checkbox"/> Kreatinine < 1.1-1.8 <input type="checkbox"/> Urinezuur < 5.5-6.42 <input type="checkbox"/> Zure fosf. (post.) < 5.8 <input type="checkbox"/> Alk. Fosfatase Value 30-125 <input type="checkbox"/> γ-GTP < 80 <input type="checkbox"/> ASAT (GOT) < 40 <input type="checkbox"/> ALAT (GPT) < 40 <input type="checkbox"/> LD (LDH) 300-450 <input type="checkbox"/> Bilirubine (nuchter) < 1.1 <input type="checkbox"/> Bilirubine (direct) < 0.2 <input type="checkbox"/> CK (CPK) < 40-200 <input type="checkbox"/> Transferrine 2.3-5.7 <input type="checkbox"/> IJzer < 12.87 <input type="checkbox"/> Cholesterol 4.1-6.4 <input type="checkbox"/> Triglyceriden 0.80-1.94 <input type="checkbox"/> Albumine 35-46 <input type="checkbox"/> TSH 0.3-3.3	HAEMATOLOGIE <input type="checkbox"/> BSE (1 uur) <input type="checkbox"/> Leucocyten APL: bevestigd <input type="checkbox"/> Differentiatie APL: bevestigd <input type="checkbox"/> Erythrocyten < 4.5-5.8 <input type="checkbox"/> Hemoglobine < 12.0-15.0 <input type="checkbox"/> Haematocriet < 0.41-0.51 <input type="checkbox"/> Thrombocyten 130-380 <input type="checkbox"/> MCV 84-102 <input type="checkbox"/> MCH 1.19-2.22 <input type="checkbox"/> MCHC 16.3-23.0	BLOEDTRANSFUSIEDIENST <input type="checkbox"/> Bloedgroep * <input type="checkbox"/> Rhesusfactor * IMMUNOLOGIE <input type="checkbox"/> Gravidatest (urine) VIROLOGIE <input type="checkbox"/> T.P.H.A. * <input type="checkbox"/> Rubella (antistoffentiter) * <input type="checkbox"/> Monolater * <input type="checkbox"/> HBsAg * FAECES-ONDERZOEK <input type="checkbox"/> Occult bloed (Haematest) 3x n3 dipet <input type="checkbox"/> Occult bloed test 1 <input type="checkbox"/> Occult bloed test 2 <input type="checkbox"/> Occult bloed test 3 <input type="checkbox"/> Vertering
ELECTROCARDIOGRAFIE <input type="checkbox"/> ECG <input type="checkbox"/> Digitalis <input type="checkbox"/> Chinidine <input type="checkbox"/> Diuretica	BACTERIOLOGIE <input type="checkbox"/> Faecesweek op <input type="checkbox"/> Salmonella/Shigella <input type="checkbox"/> Wormeieren <input type="checkbox"/> Yersinia <input type="checkbox"/> Cysten <input type="checkbox"/> Go-onderzoek Kweek (gramneg. streptoc.) <input type="checkbox"/> Urinekweek <input type="checkbox"/> Sputumkweek <input type="checkbox"/> Pusweek <input type="checkbox"/> Antiblogram	URINE-ONDERZOEK <input type="checkbox"/> Albumine <input type="checkbox"/> Urobiline <input type="checkbox"/> Bilirubine <input type="checkbox"/> Glucose (kwantitatief) <input type="checkbox"/> Glucose (kwalitatief) <input type="checkbox"/> Ketonen <input type="checkbox"/> Sediment
RADIOL. / ECHO / ENDOSCOPIE*	PATHOLOGISCHE ANATOMIE <input type="checkbox"/> Histologie <input type="checkbox"/> Anticonceptie: <input type="checkbox"/> nee <input type="checkbox"/> ja, soort: <input type="checkbox"/> Follow-up maligniteit: <input type="checkbox"/> nee <input type="checkbox"/> ja <input type="checkbox"/> Gravid: <input type="checkbox"/> nee <input type="checkbox"/> ja <input type="checkbox"/> Aspect portio: <input type="checkbox"/> Indicatie: <input type="checkbox"/> Eerder misaank: <input type="checkbox"/> nee <input type="checkbox"/> ja, nr.	PATHOLOGISCHE ANATOMIE <input type="checkbox"/> Cervix smear <input type="checkbox"/> Radiotherapie: <input type="checkbox"/> nee <input type="checkbox"/> ja <input type="checkbox"/> Latente menstruatie:

* Eventuele graviditeit vermijden

Appendix II

Two examples of feedback reports

Dr A.B.

Street name and Number

Town

Dear A.,

Within the scope of analyses of the diagnostic work-up of general practitioners in our region, I have taken a closer look at your diagnostic requests during the month of March earlier this year.

I would like to make the following remarks:

- During March you "referred" 62 patients to our diagnostic centre. In previous analyses the figures were 14-75-57-53-51 patients, while the mean numbers of referrals for all GPs in these analyses were 45-45-42-42-37.
- Regarding the filling out of the request forms I have to conclude that, as in previous analyses, this leaves much to be desired. Only 5 of all 62 forms were filled out completely. Especially the section for the findings of physical examination and the section for medication have often not been filled out. Of course I am aware of the fact that many times physical examination gives no abnormal findings and that medication is regularly unimportant, but, as stated earlier, in these cases mentioning for instance "not present" or "none" is already sufficient.

Sometimes the information of the patient is written in the wrong section (for instance, findings of physical examination mentioned in the problems/signs section). In short, I would like to ask you again to fill out the request forms properly.

- Further analysis shows how often sections were not filled out:
 - . Problems/signs 2 x
 - . Physical examination 35 x
 - . Possible diagnosis 10 x
 - . Medication 47 x
- For all 62 request forms, the "reason for request" was:
 - . Confirming the diagnosis 1 x
 - . Excluding the diagnosis 4 x
 - . Controlling a known disorder 8 x
 - . Check-up 1 x
 - . Various combinations 50 x
- Out of all available tests you requested:
 - . Clinical chemistry tests 30 x
 - . Haematological tests 34 x
 - . Serological tests 1 x
 - . Virological tests 4 x
 - . Bacteriological tests 8 x
 - . Electrocardiography 6 x
 - . Histological examination 2 x
 - . Radiological tests 20 x
 - . Ultrasonographic examination 3 x
 - . Endoscopic examination 2 x

Considering your diagnostic requests I must conclude that, apart from specifically chosen test requests, many requests have doubtful indications. You regularly ordered tests for screening while these patients have no complaints. The same happens with young patients having vague complaints and no abnormal findings in physical examination. It will not surprise you to hear that the results of these test requests give almost no extra information. In my opinion you should reduce these requests to a minimum; costs and benefit are unequal.

Please do consider the following for rational test ordering:

- * The pre-test probability for the presence of a certain disorder.
 - Combining (knowledge of, or assumptions about) the prevalence of a disorder with the results of history taking and/or physical examination, you could estimate the chance of a patient having that specific disorder.
 - If you think this chance is low and the disorder is not serious, further diagnostic work-up is not necessary. Even if the chance is high, requesting diagnostic tests is of little use. In fact only serious diseases can require further testing because then it is very useful to exclude a false positive or a false negative diagnosis.

* When you take this into account and you decide to order further diagnostic testing, you must subsequently consider which test is the best in the particular situation. This depends on:

- . The validity: some tests are appropriate to exclude a disorder, others are appropriate to confirm a disorder.
- . Some tests are invasive, others are not.
- . Some tests are very expensive, others are cheap.

Though the validity of several tests that can be useful for GPs is still unknown, you should make a choice on the basis of these factors.

* In the case of test ordering for screening purposes, note that there is a high chance of finding one or more false positive test result(s). For example, when 10 tests are requested, this chance is $\pm 40\%$! These false positive results can induce further (unnecessary) diagnostic testing.

At all instances you must consider what influence the result(s) of test(s) will have on your diagnostic or therapeutic work-up. If you think this influence is small or even absent, testing is unnecessary.

I would like to give some examples which illustrate my comments:

- In 23 cases you requested a white blood-cell count, which is, in my opinion, not indicated. Although this test is often requested by GPs for various reasons, in most instances it is of no value. An article in the Dutch Journal of Medicine, number 31 in 1988, shows the indications for the white blood-cell count.

The following examples show that you use several indications that are not mentioned in this article:

- . pt A.D. born dd-mm-yy, hepatitis because of alcoholism
- . pt A.F. born dd-mm-yy, check-up
- . pt A.G. born dd-mm-yy, recurrent urinary tract infections
- . pt A.H. born dd-mm-yy, unknown somatic disorder?
- . pt A.J. born dd-mm-yy, pruritus
- . pt A.K. born dd-mm-yy, allergy, viral infection?
- For some patients with alcoholism you requested further diagnostic tests. In these situations you mention the possible diagnosis alcoholic hepatitis.

In literature, this diagnosis is only used when alcoholic patients have jaundice, abdominal complaints (sometimes suggesting cholelithiasis) and accompanying disturbances in liver function. This disorder is, however, quite rare, therefore it is surprising that you use this diagnosis so often. Moreover, when alcoholic patients have (almost) no abdominal complaints, you cannot suspect alcoholic hepatitis; for example:

- Mrs. A.L., born 10-09-46. You requested a white blood-cell count, alk. phosphatase, Y-GTP, SGOT, SGPT and LDH. Here I have the following comments: The relevance of the WBC count is not clear to me in this situation. LDH is not a good test for estimating liver function and gives no additional information. In my opinion it is superfluous.
The test results: Alk. Phosph.: 135 U/ltr, Y-GTP: 96 U/ltr, SGOT: 25 U/ltr and SGPT: 28 U/ltr. These findings are not typical for alcoholism.
- For several patients you want a check-up, for example Mr. A.M., born 21-02-29 or Mrs. A.N., born 22-08-44. In this situation you use a set of tests: ESR, Total blood analysis, glucose, creatinine, alk. phosphatase, Y-GPT, cholesterol, electrocardiography and an X-ray of the chest. This set costs Hfl. 170. Is there any reason for such a request, particularly when a patient has no complaints? The only reason could be when it concerns high-risk patients (e.g. predisposition for cardiovascular disease).
- There is also a number of patients who underwent partially relevant and partially irrelevant testing. Some examples:
 - Mrs. A.P., born 24-03-29, suffered from low back pain. You suspected degenerative changes to be responsible for her complaints. Nevertheless, in addition to an X-ray of the lumbar spine you also requested an X-ray of the pelvis and an X-ray of the thoracic spine (costing Hfl. 105 in total).
 - Mrs. A.R., born 04-11-53, had bronchitis lasting for 2 weeks. Your request consisted of: ESR, total blood analysis, IgE, sputum culture and X-ray of the chest.
For the majority of these requests, I do not see the relevance.
 - A.S., born 17-07-82, had a furunculosis and you wanted to exclude the presence of diabetes. Besides determining glucose, you nevertheless requested also ESR, Hb and a WBC count; why?
 - Mrs. A.T., born 14-09-46, is known to have recurrent urinary tract infections; you wanted to know the cause and therefore requested ESR, WBC count, glucose, creatinine and a urine culture. For finding the cause, it would have been better to search for an anatomical factor, somewhere in the urogenital tract. Apparently, in second instance you did this regarding your request of an IVU.
- You regularly requested several tests (as a general diagnostic work-up) for patients younger than 40-45 years, with vague and unspecific complaints. In our diagnostic centre we have experienced that such a set of diagnostic tests (e.g. ESR, WBC count, total blood analysis, glucose, creatinine, Y-GPT) almost never gives any relevant result.

This is true, for example, for patient A.S. born 05-05-49, or patient A.T., born 20-09-56.

For such a group of patients it can be stated that good history-taking and a thorough physical examination are better than diagnostic testing. This is also true for patient A.V., born 07-08-55, complaining of chest pain. Your diagnosis was myalgia of thoracic muscles. Nevertheless you wanted to exclude angina pectoris and you requested an electrocardiography.

At this age atypical chest pain is unlikely to be caused by sclerosis of the coronary arteries. Moreover, the chance that this sclerosis will be detected by a simple electrocardiography without exercise is very low.

On the other hand there is also the risk that this electrocardiography gives a false positive test result with the danger of further unnecessary diagnostic testing. Altogether these are enough reasons to omit electrocardiography in such a situation.

- Finally, I would like to ask some questions about your further diagnostic (or therapeutic) work-up in several situations:

• Mr. A.W., 43 years of age, has a raised cholesterol level.

The last check-up: 8.3 mmol/ltr. What was your reaction to this situation?

• Mrs. A.Z., 25 years of age, with tiredness, nausea and vomiting, no abnormal findings in physical examination, had a raised ESR: 25 mm after 1 hour.

What did you do after this result (if you did something)?

I hope this (sixth) analysis of your test ordering behaviour is of some benefit to you. Most GPs appreciate the analyses. From you I would like to know whether you do appreciate it or not. Perhaps you have some remarks or you want to answer the questions I have raised. In that case, please do not hesitate to do so.

Yours sincerely,

P. Pop.

Dr A.Z.
Street name and Number
City

Dear A.,

Within the scope of analyses of the diagnostic work-up of general practitioners in our region, I have taken a closer look at your diagnostic requests during the month of October of this year.

I would like to give the following comments:

- During October you "referred" 31 patients to our diagnostic centre.
In previous analyses the figures were 31-14-30-29-18-18 patients, while the mean numbers of referrals for all GPs in these analyses were 45-45-42-42-37-47.

- Regarding the filling out of the request forms I can conclude that this is, although not perfect, done quite well.

Further analysis shows how often one (or more) sections of the request form have still not been filled out:

- | | |
|------------------------------------|-----|
| . Findings of physical examination | 4 x |
| . Possible diagnosis | 6 x |
| . Medication | 7 x |
| . Reason for request | 2 x |

Obviously I am aware of the fact that sometimes information is irrelevant to mention. In these cases, mentioning for instance "not present" or "none" is sufficient.

- For all 31 requests, the "reason for request" was:

. Excluding the diagnosis	1 x
. Screening	8 x
. Check-up	19 x
. Various combinations	1 x
. Not filled out	2 x

- Out of all available tests you requested:
 - . Clinical chemistry tests 8 x
 - . Haematological tests 8 x
 - . Bacteriological tests 2 x
 - . Virological tests 2 x
 - . Urine tests 1 x
 - . Electrocardiography 1 x
 - . Histological examination 14 x
 - . Radiology 3 x
 - . Ultrasonographic examination 1 x
- For a rational diagnostic request, especially mentioning a possible diagnosis is important. The corresponding pre-test probability determines whether further diagnostic tests are useful or not. In case of a high pre-test probability, further testing can be useless since (post-test) probability will be only marginally higher. The same can be true for very low pre-test probability except for the situation that you suspect the presence of a very dangerous or life-threatening disorder; then confirming the diagnosis too late can have serious consequences. In general, the information you give on the request forms is good enough, so offering a good opportunity to review the relevance of your diagnostic requests. I can conclude that your requests are normally relevant. Nevertheless there are a few situations which I would like to discuss.
 - . Miss A.B., born 10-03-69, complained of fatigue; a member of her family has mononucleosis infectiosa. Physical examination: no abnormalities. You requested: Hb, WBC count, SGOT, SGPT and Paul & Bunnell. In case you thought of a mononucleosis infection as well, it can be said that when there is no sore throat and there are no swollen lymph nodes, mononucleosis is very unlikely to be present. As a consequence, further testing seems unnecessary. Of course I am aware of the fact that many patients more or less insist on such requests. Instead of doing what these patients ask, it would be better to inform them about the often high value of history-taking and physical examination and the restricted value of diagnostic tests.
 - . A.C., born 19-09-87, suffering from diarrhoea for two months; as a result there had been weight-loss. You requested cultures for salmonella/ shigella/campylobacter/cysts. In case of diarrhoea for such a prolonged period, a bacterial cause is unlikely. However, it could be lambliasis. Also at this age, after a (short) period of diarrhoea due to an infection, frequently a prolonged period of abdominal distress, diarrhoea and flatulence can be seen, which is caused by an intolerance for lactose.

History-taking often then reveals a striking intolerance for milk or milk products. It is normally a self-limiting disorder disappearing within a few months.

As you will have noticed my comments are restricted to a few items. In the past, you always reacted to my analyses and answered my questions. Therefore I may presume that you will contact me again in the near future. I am already looking forward to it.

Yours sincerely,

P. Pop

Test reports

Differential count and monospot from the end of the first week of the disease.

Appendix III

Guidelines used in the Maastricht region

- Electrocardiography
- TSH as a thyroid function test
- Faecal cultures for diarrhoea
- The diagnostic work-up for anaemias
- White blood-cell counts and differential counts
- Excessive bleeding
- Liver tests
- The diagnostic work-up for infectious mononucleosis
- Urine cultures
- Renal function tests
- The diagnostic work-up for rheumatoid arthritis
- Uric acid and gout
- The diagnostic work-up for tiredness/general discomfort
- The diagnostic work-up for allergic disorders
- Ultrasonography and symptomatic chole(dochol)lithiasis
- Ultrasonography of liver and biliary tract
- Radiologic imaging of the sinuses
- Radiologic imaging of the chest
- Radiologic imaging for low back pain
- Radiologic imaging of the vertebral column, joints and major bones
- Diagnostic imaging of renal stones
- Endoscopy
- Cervical smears
- Cholesterol determinations

Appendix IV

An example of the guidelines

The diagnostic work-up for infectious mononucleosis

Goal

To confirm or exclude infectious mononucleosis.

Indications

Complaints and signs of fever, sore throat, enlarged cervical lymph nodes, oral petechiae (palate), splenomegaly.

Comments on the indications

In children infectious mononucleosis can be present without complaints or signs. The disease is most typically seen in adolescents and young adults. Infectious mononucleosis is seen only sporadically in patients older than 30 years; it is not seen in patients older than 40 years. The findings of physical examination are important. When fever, enlarged cervical lymph nodes, splenomegaly and oral petechiae are absent, infectious mononucleosis is not present and laboratory tests are not necessary. The presence of infectious mononucleosis is unlikely when enlarged cervical lymph nodes and splenomegaly are absent. Patients who complain of sore throat often have also one of the following complaints or signs: fever, enlarged cervical lymph nodes, splenomegaly, oral petechiae of the palate. Then the possibility of infectious mononucleosis must be considered. Splenomegaly and petechiae have a high specificity for infectious mononucleosis, whereas fever and enlarged cervical lymph nodes do not have a high specificity. Infectious mononucleosis is a self-limiting disease, characterized by a post-infectious period of tiredness and apathy. There is no correlation between the severity of complaints and the extent of abnormality of laboratory test results (level of Paul-Bunnell, level of liver tests). Chronic hepatitis nor cirrhosis do not occur. Consequently it is not necessary to check or repeat laboratory tests.

Test requests

Differential count and monospot from the end of the first week of the disease.

Interpretation

Typical changes in the differential count of leucocytes (more than 50% lymphocytes, of which more than 10 atypical) can be seen as soon as lymph nodes are enlarged at the end of the first week. These haematological abnormalities are at their maximum in the second or third week and are present for one or two months. A comparable lymphocytosis can be seen cytomegalovirus infections and infectious hepatitis.

In 75% of all patients with infectious mononucleosis antibodies can be found through Paul-Bunnell or monospot at the end of the first week. After three weeks this is in virtually 100% of all patients. In 8-12 weeks after the onset of the disease they start to disappear, but they can remain present for about 6 months. A level higher than 1:32 can be considered to be positive.

Comments on the interpretation

Studies have shown that the differential count, Paul-Bunnell and monospot have a high discriminating power. Erythrocyte Sedimentation Rate (ESR), leucocyte count and liver tests have a low discriminating power and thus are of little value in the diagnostic work-up of infectious mononucleosis.

Test characteristics of the differential count, Paul-Bunnell and monospot are given in the following table.

Sensitivity, specificity and likelihood ratios

Test	Definition of a positive test result		Sens. %	Spec. %	LR ⁺	LR ⁻
Differential count	> 50% lymphocytes	end of first week	82	98	41	0.18
	of which > 10% atypical	highest level:	90	98	45	0.10
Paul-Bunnell	level > 1:224	end of first week:	85	99	85	0.15
		highest level:	97	99	97	0.03
Monospot	level > 1:224	end of first week:	85	98	43	0.15
		highest level:	96	98	48	0.04

Normal values and costs

Test	Normal values	Charges	
		HS*	PI**
Differential count	lymphocytes 20-45%	f 3,-	3,95
Paul-Bunnell	$\geq 1:32$	f 7,20	13,95
Monospot	negative	f 7,20	13,95

* HS = Health Insurance Service
 ** PI = Private Insurance

Appendix V

The costs of feedback

There are two main possibilities for providing feedback.

In the first situation feedback is provided by an individual peer, not involved in any diagnostic centre. In this situation, there is only little extra administrative support needed since activities are solely restricted to the provision of feedback. It requires mostly the development of new request forms (containing clinical information), the collection of request forms and data analysis, the review of request data and the processing of feedback reports. Based on the experiences of the Diagnostic Coordinating Centre Maastricht, reviewing all the requests from one month of ± 85 general practitioners will take approximately 2 hours of time per general practitioner. This is done twice each year. Data collection and data analysis will take approximately 28 hours per week, while for processing feedback reports on average 20 hours per week are needed. Together this costs f 95000 per year (\$ 50000). Materials cost approximately f 9500 yearly (\$ 5000).

If feedback is provided only occasionally a peer reviewer and administrative personal will be less familiar with the procedures involved and as a result the feedback procedure may be more time-consuming. Feedback will then cost at most (at an estimate) f 155000 (\$ 85000) per year.

In the second, optimal, situation there is one centre where all diagnostic requests are coordinated as a part of routine health care and thus all request forms are automatically collected. Far more activities than the feedback alone are possible then. In this situation, apart from the peer reviewer and the administrative support mentioned above, additional administrative support is needed for a diversity of health care activities. Obviously, this requires also higher expenditures for materials, interest and depreciation. Especially more administrative support is needed, with a minimum of 60 hours per week. Costs for all staff then are f 180000 per year (\$ 100000). Costs of materials, etcetera, are set at f 35000 per year (\$ 19000). The expenditure for providing feedback twice a year (see above) is then a part of this total amount of costs.

Improving test ordering in general practice

Ron Winkens

Can routine individual feedback improve the quality of test ordering behaviour and reduce the number of unnecessary requests? If it does, to what extent and for how long is test ordering behaviour changed? These questions are dealt with in this book. Feedback causes change, at times considerable, the most important being quality improvement. Change persists when the feedback is repeated.

The longer it is continued, the higher the return.

Other relevant factors, like referral behaviour and doctors' characteristics, are also considered. An example is given how to deal with the unbridled growth in the volume of diagnostic testing.

Ron Winkens studied medicine at the University of Limburg, Maastricht, the Netherlands. He works as a researcher at the Diagnostic Coordinating Centre Maastricht and also as a general practitioner. His fields of interest are quality assurance on diagnostic testing and the use of computers in general practice. He sees the combination of these two topics as a continuing challenge.

ISBN 90-5170-258-2



9 789051 702583

NUGI 741